

## **Tideless estuaries in brackish seas as possible freshwater-marine transition zones for bacteria**

Golebiewski, Marcin; Calkiewicz, Joanna; Creer, Simon; Piwosz, Kasia

### **Environmental Microbiology Reports**

DOI:

[10.1111/1758-2229.12509](https://doi.org/10.1111/1758-2229.12509)

Published: 01/04/2017

Peer reviewed version

[Cyswllt i'r cyhoeddiad / Link to publication](#)

*Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):*

Golebiewski, M., Calkiewicz, J., Creer, S., & Piwosz, K. (2017). Tideless estuaries in brackish seas as possible freshwater-marine transition zones for bacteria: the case study of the Vistula river estuary. *Environmental Microbiology Reports*, 9(2), 129-143. <https://doi.org/10.1111/1758-2229.12509>

#### **Hawliau Cyffredinol / General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**Tideless estuaries in brackish seas as a possible freshwater-marine transition zones for bacteria – the case study of the Vistula river estuary**

Marcin Gołębiewski<sup>1,2</sup>, Joanna Całkiewicz<sup>3</sup>, Simon Creer<sup>4</sup>, Kasia Piwosz<sup>3,5\*</sup>

1 – Chair of Plant Physiology and Biotechnology, Nicolaus Copernicus University, Lwowska 1, 87-100 Toruń, Poland

2 – Centre for Modern Interdisciplinary Research, Nicolaus Copernicus University, Wileńska 4, 87-100 Toruń, Poland

3 – National Marine Fisheries Research Institute, Kołłątaja 1, 81-332 Gdynia, Poland

4 – School of Biological Sciences, Bangor University, Bangor, Gwynedd, LL57 2UW, UK

5 – Center Algatech, Institute of Microbiology Czech Academy of Sciences, Novohradská 237, 37981 Třeboň, Czech Republic

**\*Corresponding author:**

Name: Kasia Piwosz

Address: Center Algatech, Institute of Microbiology Czech Academy of Sciences, Novohradská 237, 37981 Třeboň, Czech Republic

Telephone: +420 384 340 482

Fax: +420 384 340 415

E-mail: piwosz@alga.cz

**Running title:** Bacterial freshwater-marine transition in the Baltic Sea

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1758-2229.12509

## Originality-Significance Statement

This is the first publication addressing the hypothesis that estuaries in brackish seas are likely environments where freshwater bacteria could adapt to marine conditions. We showed that a number of freshwater bacterial phylotypes thrived in brackish waters at salinities of approximately 7. Our results suggested selection acting on a pool of highly similar subOTUs (OTUs at 1% dissimilarity level constructed from a given 3% dissimilarity OTU, likely strains of the same species) as a plausible mechanism enabling freshwater-marine transitions of bacterial taxa.

## Summary

Most bacteria are found either in marine or fresh waters and transitions between the two habitats are rare, even though freshwater and marine bacteria co-occur in brackish habitats. Estuaries in brackish, tideless seas could be habitats where the transition of freshwater phylotypes to marine conditions occurs. We tested this hypothesis in the Gulf of Gdańsk (Baltic Sea) by comparing bacterial communities from different zones of the estuary, via pyrosequencing of 16S rRNA amplicons. We predicted the existence of a core microbiome (CM, a set of abundant OTUs present in all samples) comprising OTUs consisting of populations specific for particular zones of the estuary. The CMs for the entire studied period consisted of only eight OTUs, and this number was even lower for specific seasons: five in spring, two in summer, and one in autumn and winter. Six of the CM OTUs, and another 21 of the 50 most abundant OTUs consisted of zone-specific populations, plausibly representing micro-evolutionary forces. The presence of up to 15% of freshwater phylotypes from the Vistula River in the brackish Gulf of Gdańsk supported our hypothesis, but high dissimilarity between the bacterial communities suggested that freshwater-marine transitions are rare even in tideless estuaries in brackish seas.

## Introduction

Marine and freshwater bacterial communities are distinct, even differing conspicuously at the phylum level (Logares *et al.*, 2009). Even in estuaries, where they could theoretically co-exist, freshwater lineages rapidly diminish within increasingly saline waters, while marine groups are absent from the freshwater regions of an estuary (Crump *et al.*, 2004). The low frequency of freshwater-marine transitions (i.e. adaptation of freshwater organisms to increased salinity) indicates that this process might be driven by rapid rather than gradual adaptations (Logares *et al.*, 2009; 2010). However, estuaries in brackish, tideless seas could represent habitats where transitions of freshwater phylotypes to marine conditions is facilitated, compared to typical estuaries. In such areas, mixing of riverine and marine water masses is slower due to the lack of tide currents, and the overall salinity gradient is mild. Therefore, brackish seas and estuaries offer a unique opportunity to study coexistence of marine and freshwater bacteria, formation of unique brackish communities and occurrence of freshwater-marine transitions (Riemann *et al.*, 2008; Herlemann *et al.*, 2011).

A typical example of a brackish estuary is the Gulf of Gdańsk (southern Baltic Sea). The salinity of surface waters that varies between 7 and 8 in the proper Baltic Sea is lowered to 6-7 in the Gulf of Gdańsk by freshwater runoff from the River Vistula (Kowalkowski *et al.*, 2012). The freshwater inflow affects the composition and activity of bacterial communities in the open part of the Gulf of Gdańsk (Ameryk *et al.*, 2005; 2014). Typical freshwater phylotypes, e.g. R-BT lineage (*Limnohabitans*) of Betaproteobacteria, Ac1 Actinobacteria, LD12 Alphaproteobacteria, are an active component of bacterial communities in coastal waters of the Gulf of Gdańsk, showing pronounced seasonal dynamics (Piwosz *et al.*, 2013). This suggests that freshwater bacterial phylotypes from the Vistula River may adapt to brackish conditions and contribute to brackish communities. Based on these observations, we hypothesized that freshwater-marine transitions might be facilitated in brackish estuaries.

To address this hypothesis, we studied the composition of active bacterial communities in three zones of the Vistula River estuary: freshwater (salinity (S) < 0.5), mixing zone (S~3.5) and brackish

(S~7) by high-throughput sequencing of V3-V4 fragments of bacterial 16S rRNA with environmental rRNA as a template. We expected the following:

1. Some bacterial phylotypes are present in all zones of the Vistula estuary (the river, mixing zone and brackish waters of Gulf of Gdańsk) – in other words, a delineation of the core estuary microbiome (a set of OTUs abundantly present in all zones of the estuary) is possible, and consequently, the influence of the river on the biodiversity of the mixing zone and Gulf of Gdansk can be observed;
2. Core microbiome phylotypes defined at the 97% similarity cutoff (might be regarded as 'species') consist of 99% similarity OTUs (subphylotypes, likely representing strains) specific (i.e. most abundant in) for particular zones of the estuary (salinity) and season (temperature).

Additionally, we estimated the phylogenetic diversity of bacterial communities in the three zones of the estuary and predicted that the mixing zone, which is a transition zone between freshwater and marine ecosystems (ecotone), would harbour the greatest bacterial diversity, whilst the Gulf of Gdańsk would possess the lowest due to the least trophic conditions.

## Results

### *Environmental characterization of the sampling stations*

Concentration of nutrients, chlorophyll-*a* and bacteria were the highest at the freshwater site and the lowest at the brackish site (Table 1). Concentrations of nutrients were the highest in January, and the lowest in June at all sites, while concentrations of chlorophyll-*a* showed the opposite pattern. Bacteria concentrations reached maximum in summer and minimum in winter at the freshwater site and in the mixing zone, while at the brackish site, the maximum abundance was found in autumn and minimum in spring (Table 1). The shares of the Vistula waters in the mixing zone ranged from 0.56 (October 2011) to 0.70 (January 2012, Table 2), and concentrations of nutrients, chlorophyll-*a* and bacteria were intermediate in the mixing zone (Table 1). The measured concentrations agreed

with the theoretical values computed from the shares of fresh- and brackish waters, except for July 2011, when the expected concentrations were up to 22 times higher than those measured (Table 2).

Principal Component analysis clustered the samples according to their origin (salinity) and season (temperature, Fig. 1). The first principal component that explained 52.9 % of the variance in the environmental variables, correlated positively with salinity and temperature, and negatively with concentrations of dissolved silica, total phosphorus and total nitrogen. The second principal component explained 38.8 % of the variance, and correlated positively with chlorophyll-*a*, and negatively with salinity as well as concentration of total nitrogen.

Differences between maximum and minimum values of abiotic factors were larger for the freshwater site and the mixing zone than for the brackish site (PERMDISP analysis: betadisper and anova,  $p < 0.01$ ), indicating that environmental conditions were more stable in the Gulf of Gdańsk than in the Vistula River and the mixing zone.

#### *Sequencing statistics*

1 280 729 raw reads (405 904 unique) were generated and have been deposited in the NCBI short read archive (SRA) database under accession number **SRP064705**.

The number of the unique sequences was reduced to 143 009 by denoising (correction of presumable PCR and pyrosequencing errors) and 547 739 sequences (45 907 unique) covered the desired region of the SILVA alignment (6 500-22 500).

As chimera detection algorithms differ significantly, we employed a multi-step chimera detection procedure. A combination of Uchime, Perseus and chimera.slayer with reference database worked best, removing together 55 917 of putative chimera sequences (35 737 (10 520 unique) by UCHIME, 6 617 (1 521 unique) by Perseus, and 13 563 (452 unique) by chimera.slayer). Moreover, 151 951 reads affiliated with chloroplasts and 57 522 singletons and doubletons were removed, leaving 282 349 sequences for further analyses. The error rate was estimated to be  $6.28 \times 10^{-5}$  errors/base.

### *Bacterial communities in different salinity zones of the Vistula estuary*

6 139 OTUs were identified at the 97% similarity level (Supplementary Table 1), affiliated with 32 bacterial phyla. Rarefaction analysis indicated that the samples were moderately sampled (Supplementary Fig. 1).

Species richness was consistently lower in brackish samples, than in freshwater and mixing zone ones (Fig. 2A and B), and evenness as well as diversity were similar in all sample types in all seasons but summer, when they were the lowest in the brackish samples and the highest in the freshwater (Fig. 2C and D, ANOVA with the Tukey's test,  $p < 0.01$ ).

Almost all reads were classified to the phylum level: unclassified reads comprised  $< 1\%$  of all sequences. However, from almost 7% up to 46% of reads per sample could not be classified at the family level, and 13-38% of reads belonged to rare families. The majority of reads were classified to the genus level (59-91.5% per sample), and 32-57% of reads belonged to rare taxa (comprising altogether below 1% of all reads). Consistently more reads were classified in brackish samples from the Gulf of Gdansk than in the freshwater and mixing zones.

Proteobacteria, Actinobacteria, Cyanobacteria and Bacteroidetes comprised together between 77% to over 95% of reads and were present at all sites in all seasons (Fig. 3A). The bacterial communities appeared to be similar between the sites at the level of phylum, but analyses at lower taxonomical levels revealed conspicuous differences between the sites and seasons (Fig. 3B-D).

Alphaproteobacterial reads (Rhodobacteraceae and Sphingomonadaceae) were most abundant in the Vistula river libraries at each sampling time (Fig. 3B-D). Actinobacterial sequences (Sporichthyaceae hgcI clade - acI Actinobacteria) were abundant in spring and autumn, whilst Betaproteobacterial (*Limnohabitans*) reads prevailed in winter at the freshwater site. Contributions of other bacterial groups to the active community in the Vistula River were sporadic: Flavobacteria appeared in spring and summer, and Planctomycetes, Verrucomicrobia and Cyanobacteria in summer (Fig. 3B-D).

Active bacterial communities in the mixing zone were very similar to those in the Vistula River, with dominance of Alphaproteobacterial reads (Rhodobacteraceae and Sphingomonadaceae) in each season, a high contribution of Actinobacterial reads (Sporichthyaceae hgcI clade - acI Actinobacteria) in spring and autumn, and of *Limnohabitans*-derived sequences in winter (Fig. 3A-D). However, an increased contribution of *Synechococcus* (Cyanobacteria Subdivision I Family I) and *Anabaena* (Cyanobacteria Subdivision I Family I) sequences was observed in summer, the two taxa that dominated the libraries from the brackish site.

Alphaproteobacteria (Rhodobacteraceae) were also a constant component of bacterial communities in brackish waters (Fig. 3A-D). Actinobacteria (other than Sporichthyaceae hgcI clade) also substantially contributed to the active bacterial community at the brackish site except during summer. Acidimicrobiaceae, Planctomycetes and Verrucomicrobia contributed in autumn and winter, while Cyanobacteria (*Synechococcus* and *Anabaena*) conspicuously dominated in summer (Fig. 3A-D).

The first NMDS axis discriminated samples according to salinity, and the second one according to season (Fig. 4A). Active bacterial communities in the brackish waters were significantly different from communities in the freshwater and the mixing zone (AMOVA and ANOSIM,  $p < 0.01$ , Supplementary Table 2), but the latter two were similar.

The most abundant OTUs differentiated brackish communities from those from the freshwater and mixing zone ones. OTU1 (Sporichthyaceae), OTU3 (Sphingomonadaceae), OTU4 (*Limnohabitans*), OTU5 (Sporichthyaceae) and OTU6 (Rhodobacteraceae) were highly abundant in freshwater and mixing zone libraries, while OTU2 (*Synechococcus*), OTU8 (BAL58 marine group), OTU10 (Sporichthyaceae), OTU14 (*Snowella*) and OTU15 (PeM15) were characteristic for the brackish waters (Fig. 4B).

#### *Core microbiome of the Vistula River estuary*

We defined the core microbiome as a set of OTUs that contributed at least 1% of reads to each sample. No OTUs present at all sites in all seasons above the abundance threshold were found, but



seasonal core microbiomes were defined. Eight OTUs contributed to the so defined core microbiomes of the Vistula River estuary. OTUs were classified as 'freshwater', 'mixing zone', 'brackish' or 'variable' according to their peak contribution to the libraries (see Experimental procedures). The spring microbiome consisted of five OTUs: freshwater OTU1 (hgcl Sporichthyaceae, Actinobacteria), OTU3 (unclassified Sphingomonadaceae, Alphaproteobacteria) and OTU5 (unclassified Sporichthyaceae, Actinobacteria), and mixing zone OTU6 (unclassified Rhodobacteraceae, Alphaproteobacteria) and OTU20 (GKS98 freshwater group, Betaproteobacteria). The summer core microbiome contained two OTUs: a brackish OTU7 (*Anabaena*, Cyanobacteria) and a mixing zone OTU19 (*Hyphomonas*, Alphaproteobacteria). Autumn and winter core microbiomes comprised only a single freshwater OTU each: OTU25 (hgcl Sporichthyaceae Actinobacteria) and OTU3 (unclassified Sphingomonadaceae, Alphaproteobacteria), respectively.

#### *Influence of the Vistula River and the Gulf of Gdansk waters on bacterial communities in the mixing zone*

Active bacterial communities in the mixing zone samples consisted predominantly of freshwater and mixing zone phylotypes (Fig. 5). The largest number of freshwater OTUs in the mixing zone was observed in summer (Fig. 5A). However, the numbers of reads coming from these OTUs were moderate, suggesting low activity and/or cells numbers (Fig. 5B). The greatest number of reads of freshwater OTUs was found in spring samples (Fig. 5A and B), indicating possible high activity and/or number of freshwater bacteria in the mixing zone at that time. The smallest contribution of the freshwater OTUs was observed in winter.

The contribution of OTUs that were characteristic for the mixing zone followed the share of the Vistula waters, but such a relationship was not observed for the number of brackish and freshwater OTUs (Fig. 5B). Nevertheless, among the 100 most abundant OTUs, the abundance of 13 taxa was significantly greater in the mixing zone derived libraries than expected from passive mixing alone,

assuming that the Vistula River and the Gulf of Gdańsk were the only sources of bacterial cells (Table 3, Supplementary Table 3).

The greatest fraction of brackish OTUs in the mixing zone was found in autumn (Fig. 5A), while the number of reads coming from the brackish OTUs were the highest in summer (Fig. 5B). An ANOVA analysis indicated that abundance of reads originating from Alphaproteobacteria was higher than expected from freshwater and brackish water shares. The differential abundance, mainly derived from differences in the abundance sequences coming from the Caulobacterales order, was higher than expected, pointing to a possibility that they contributed to this apparent increase in activity of bacteria during summer (Table 3, Supplementary Table 3). Interestingly, neither the high fraction of brackish OTUs nor the fraction of reads coming from them coincided with low contributions of freshwater OTUs to the mixing zone.

*Influence of Vistula on Gulf of Gdańsk waters – certain freshwater bacterial taxa are active in brackish waters*

We were interested in the identification of freshwater taxa able to thrive in brackish Baltic waters. Freshwater OTUs represented by at least four reads in Baltic libraries were deemed as able to live in Baltic. The sets of such OTUs were different in each season (Table 4).

Contribution of freshwater OTUs to the Gulf of Gdańsk waters was close to 15% in spring, summer and autumn, and < 8% in winter (Fig. 5C). In terms of the number of reads the influence was even lower, reaching maximum of 9.4% in spring (Fig. 5D). Interestingly, although the movement of the water masses is usually unidirectional from the river to the sea, a brackish OTU2 (*Synechococcus*) was found in minor quantities (0.04-0.64%) at the freshwater site.

*Certain OTUs consist of many more resolved OTUs that display ecologically specific distributions in relation to sites and seasons*

To check whether the OTUs consisted of zone and season-specific subphylotypes, 1% dissimilarity OTUs (hereafter referred to as 'subOTUs') were constructed for 50 of the most abundant OTUs (OTU1 to OTU50) and phylogenetic trees of these OTUs annotated with their peak abundance were

constructed. Twenty three OTUs consisted of only one major subOTU or of subOTUs that displayed similar patterns of spatial and temporal occurrence in our samples. Twenty seven OTUs, including six core ones, consisted of subOTUs whose abundances differed across sites and seasons, among them were 17 OTUs that changed their pattern of occurrence over the year (Supplementary Fig. 2). Core microbiome OTU20 (GKS98 freshwater group of *Alcaligenaceae*) was an example of an OTU consisting of multiple 99% similarity OTUs specific for different zones of the estuary. It was divided into five abundant subOTUs: two most closely related (subOTUs 4 and 5) were specific for the brackish waters, while subOTUs 1, 2 and 3 were specific for freshwater/mixing zone (Fig. 6A). The combined distribution patterns of these subOTUs resulted in a peak abundance in the mixing zone in spring and in freshwater in other seasons for OTU20. An example of an OTU consisting of subOTUs with similar patterns of abundance was OTU18 (Family I, Subsection III of Cyanobacteria, Fig. 6B). All of its three abundant subOTUs were specific for freshwater/mixing zone, but differed in time of their peak abundance: reads coming from subOTU1 were most numerous in autumn and winter, from subOTU2 in winter and spring, and from subOTU3 in summer. On the other hand, the core microbiome OTU19 (*Hyphomonas*) consisted of one abundant subOTU which displayed a variable pattern characteristic for the whole OTU19 (Supplementary Fig. 2).

## Discussion

Here, we showed a number of freshwater phylotypes (OTUs) from the Vistula river to thrive in the brackish waters of the Gulf of Gdańsk. We also showed that some of these phylotypes consist of subphylotypes displaying ecological differentiation across different habitats: freshwater, mixing zone or brackish. Collectively, these findings support the hypothesis that the freshwater-marine transition of certain bacterial taxa might be facilitated in tideless estuaries in brackish seas.

We focused on active bacteria in the Vistula river estuary by sequencing fragments of transcribed 16S rRNA (and not rRNA genes). The rRNA:rDNA ratio is often used to monitor

changes in bacterial activity (Campbell *et al.*, 2013), but this approach has been recently disputed, as the relationship between non-growth activity and rRNA concentration is currently not known, moreover dormant cells can contain high number of ribosomes (Blazewicz *et al.*, 2013).

Nevertheless, as we were concerned that a large fraction of DNA in the mixing zone could have originated from dead bacteria, we focused on rRNA, and in further analysis we avoided comparing the abundance of reads between different OTUs, but instead we compared the abundance of reads of specific OTUs between different samples (Ibarbalz *et al.*, 2014). It must be acknowledged that a number of reads originating from a given phylotype in our study is not a true measure of bacterial activity, as high read numbers may be found for low activity organisms if they are sufficiently abundant, and conversely, low number of reads would be obtained even for active but rare organism. Nevertheless, numbers of reads do provide indirect information and an appropriate dependent variable that likely represents changes in activity and/or abundance of specific phylotypes in the different zones of the estuary.

The temporal dynamics of specific bacterial phylotypes is high in the Gulf of Gdańsk (Piwosz *et al.*, 2013), and more frequent sampling could have resulted in the detection of transient phylotypes and, thus, a larger core microbiome of the Vistula estuary. Nevertheless, we recovered most of the taxa reported previously in freshwaters (Newton *et al.*, 2011), estuaries (Campbell *et al.*, 2013) and the Baltic Sea (Anderson *et al.*, 2010; Herlemann *et al.*, 2011). The similarity of our results to those reported by other groups, and the fact that the environmental conditions were typical for the Vistula estuary and Gulf of Gdańsk (Ameryk *et al.*, 2005; 2014; Wielgat-Rychert *et al.*, 2013), suggests that we captured most of the phylogenetic diversity of the Vistula River estuary throughout the year. Therefore, acknowledging the limited number of temporal observations here, we predict that more frequent sampling would not contradict our conclusions, but would rather support them more strongly.

*Influence of the Vistula River on bacterial diversity and active bacterial communities in the Gulf of Gdańsk*

The relation between salinity and benthic macro species diversity in the Baltic Sea forms a parabola-shaped curve with a minimum (termed horohalinicum) at salinity 5-8 (Remane, 1934). We expected the mixing zone, an ecotone between fresh- and brackish waters, to bear the greatest diversity, and the brackish site, least productive and within horohalinicum, to be the least diverse. However the diversity in the mixing zone was not significantly higher, but comparable to that in the Vistula River (Fig. 3). Our findings support the hypothesis that the Remane curve is inappropriate for planktonic microorganisms (Herlemann *et al.*, 2011; Telesh *et al.*, 2011a; 2011b).

The Vistula River strongly affected bacterial communities in the mixing zone, as can be deduced from the overall high similarity of the freshwater and mixing zone communities (Fig. 4, Supplementary Table 2). The number of freshwater phylotypes in the mixing zone exceeded the number of brackish phylotypes in all seasons but summer, indicating that freshwater bacteria (particularly Alphaproteobacteria) may have adapted more readily to the conditions in the mixing zone. This agrees with observations that freshwater bacteria can be abundant at salinity of 3 to 4 (Herlemann *et al.*, 2011; Campbell *et al.*, 2013). The brackish phylotypes were highly active in the mixing zone only in summer (Fig. 5). At that time, the measured concentrations of bacteria and nutrients (but not of chlorophyll-*a*) substantially differed from the theoretical values, calculated from shares of the Gulf of Gdańsk and the Vistula waters (Witek *et al.*, 2003; Wielgat-Rychert *et al.*, 2013) (Table 2). This disagreement, together with the fact that the increase of bacterial cells numbers did not explain the observed reduction of nutrients concentrations and the lack of increase in chlorophyll-*a* levels (indicating that phytoplankton photosynthetic activity did not increased) suggested that the substrates might have been depleted due to the increased activity of bacteria in the mixing zone (Ameryk *et al.*, 2005; 2014; Wielgat-Rychert *et al.*, 2013). This view is substantiated by an observation that bacterial activity was stimulated under decreased salinity and increased resources conditions in an experimental transplantation of bacterial communities between the Baltic Proper and the Bothnian Sea (Lindh *et al.*, 2015). Phylotypes whose numbers of reads in the mixing zone libraries were greater than expected from fresh- and brackish water shares

(Supplementary Table 3) might have been responsible for this apparent increased activity.

Phylotypes demonstrating this greater activity belonged mainly to Alphaproteobacterial genera such as *Brevundimonas*, *Hyphomonas*, or *Roseibacterium* (Table 3).

The effect of the Vistula River on active bacterial communities was less evident in the brackish waters of the Gulf of Gdańsk (Table 4; Figs. 3 and 4). Brackish assemblages consisted of bacteria typical for the Baltic Sea: Actinobacteria, Alphaproteobacteria and Cyanobacteria (Anderson *et al.*, 2010; Herlemann *et al.*, 2011). A notable difference between our and previous research was the virtual absence of the reads from Verrucomicrobia and SAR11 from our libraries. As we used the primers designed by Herlemann *et al.* (2011), it seems plausible that the difference resulted from low activity of these bacteria in the Gulf of Gdańsk, making them hard to detect in an RNA-based study. A lack of activity, as opposed to putative primer bias, is supported by the fact that

Verrucomicrobial reads were detected in the freshwater libraries, but not at the brackish site in summer (Fig. 4). Moreover, the low metabolic activity of the marine SAR11 clade in the Gulf of Gdańsk has been observed previously and was attributed to decreased salinity and increased trophic conditions (Piwosz *et al.*, 2013).

#### *Are tideless estuaries in brackish seas zones of facilitated freshwater-marine transition of bacterial phylotypes?*

The core microbiome is a concept aimed at showing the most important organisms in a set of samples (Turnbaugh *et al.*, 2009). As low-frequency OTUs might represent organisms with very low activity or rare in a given zone of the estuary, we chose the 1% abundance threshold, resulting in a low number of core OTUs (8). Lowering the threshold would broaden the core set, up to the limit of shared OTUs number (i.e. at least one read in each sample), which would give the core microbiome consisting of 79-211 (7.91-13.98%) OTUs, depending on the season. Nevertheless, the delineation of season-specific core microbiomes in the Vistula River estuary was unexpected, as it has been impossible to achieve for sets of samples with much more similar physicochemical

properties e.g. a set of human gut samples (Turnbaugh *et al.*, 2009) or soils (Zarraonaindia *et al.*, 2015).

The presence of freshwater phylotypes in the Gulf of Gdańsk, combined with the existence of the core microbiome in the Vistula estuary, supported our hypothesis that tideless estuaries in brackish seas may facilitate freshwater-marine transition of microorganisms. In addition, most core OTUs were freshwater phylotypes, which supports the view that adaptation of freshwater bacteria to brackish conditions (salinity 6-7) is easier than adaptation of brackish ones to freshwater conditions (Lindh *et al.*, 2015). On the other hand, the number of freshwater phylotypes was high in the mixing zone but low in the open waters of the Gulf of Gdańsk (Fig. 5). The paucity of freshwater taxa in brackish waters suggests that adaptation to salinity between 4 and 7 may be more difficult than adaptation to salinity <4. The possible causes may include greater difficulty of physiological adaptations, but also ecological processes, such as competition or grazing pressure (Ruiz *et al.*, 1998).

Six of eight core OTUs consisted of 1% dissimilarity OTUs (subOTUs) with contrasting spatial and/or temporal distribution patterns. Our approach to define subOTUs is operationally exchangeable with a procedure termed 'oligotyping'. In oligotyping, Shannon's information entropy is calculated for all positions in the alignment, which allows the discrimination of true variations (having high entropy) from random ones. The methodology proved to be useful for non-denoised datasets, as it allowed for detection of true variants in noisy data (Eren *et al.*, 2013). As we employed rigorous noise removal approach, we consider the subOTUs to represent true variants of 16S rRNA sequences.

In the core microbiome three of the OTUs that consisted of site specific subOTUs belonged to Actinobacteria (Sporichthyaceae), and three to Alphaproteobacteria (members of Sphingomonadaceae, Rhodobacteraceae and the *Hyphomonas* genus). Sporichthyaceae are ubiquitous in freshwaters (Newton *et al.*, 2011), and were abundant in oligosaline arctic lakes (Theroux *et al.*, 2012) as well as in the Baltic Sea (Anderson *et al.*, 2010; Piwosz *et al.*, 2013), and



so it is likely that more bacteria of this family tolerate salinity of up to 3-4 PSU, and, to a lesser extent 7-8 PSU. Similarly, typically marine members of the *Hyphomonas* genus (Li *et al.*, 2014) appear to be able to tolerate mixing zone and freshwater conditions. Both marine (e.g. Zhang *et al.*, 2015; Zhao *et al.*, 2016) and freshwater (e.g. Chen *et al.*, 2013; Park *et al.*, 2014) bacteria are known among Sphingomonadaceae and Rhodobacteraceae, demonstrating their ability to cope with varying salinity, thus their presence in the estuarine core microbiome is plausible.

Selection acting on a pool of similar strains could have been the mechanism responsible for adaptation to different salinities. Bacteria may thrive in changing environmental conditions due to physiological responses at various levels (transcription, translation, post-translational modifications, and allosteric regulation). Such responses may result from phenotypic plasticity or from adaptations involving genetic changes (single nucleotide changes, horizontal gene transfer, large genome rearrangements (Ryan, 1952; reviewed in Ryall *et al.*, 2012) that could be manifested also in 16S rRNA genes. In such case, the 16S rRNA gene variants could be linked with variants of functional traits. In our study we found that over half of the 50 most abundantly represented bacterial phylotypes consisted of subphylotypes (subOTUs) preferentially found at a particular site or in a particular season, which may suggest that adaptations are frequent mechanism leading to freshwater-marine transition. However, it should be borne in mind that the importance of phenotypic plasticity driven by changes in gene expression pattern could not have been directly assessed in this study.

Collectively, our results support the hypothesis that freshwater-marine transitions might be accomplished in tideless estuaries in brackish seas. Nevertheless, the low similarity between bacterial communities from the Vistula River and the Gulf of Gdańsk indicated that such events were rare. We propose that the mechanism responsible for adaptation to different salinities is likely to be microevolutionary forces imposing selection on ecologically heterogeneous, but taxonomically closely related bacteria from a pool of similar strains. This hypothesis needs further investigations, and may be an avenue for future research.



## Experimental procedures

### *Sampling*

Samples were collected in July and October 2011, as well as in January and April 2012, at three sites in the Vistula Estuary (Fig. 7). Salinity and temperature were measured *in situ* with a Cast Away CTD probe (SonTec YSI Inc, USA) and the exact positions for the mixing zone and brackish sites were decided based on the salinity values.

Twenty-five litres (L) of surface water were collected in triplicates using a Niskin bottle and approximately 20 L per sample was filtered through a 20 µm mesh plankton net into acid and ethanol-sterilized containers, washed thoroughly with the sampled water. The sampled water was subsequently used for RNA extraction and for cell counts. Five litres per sample of the unfiltered water was stored in light-proof containers for downstream nutrient and chlorophyll-a analysis.

### *Calculation of the proportion of the Vistula and Gulf of Gdansk waters in the mixing zone*

The proportion of fresh and brackish water masses in the mixing zone were calculated from the following formulae:

$$1. f_r = (S_m - S_b) / (S_r - S_b),$$

$$2. f_b = 1 - f_r,$$

where  $f_r$  denotes fraction of freshwater,  $f_b$  denotes fraction of brackish water, and  $S_r$ ,  $S_m$  and  $S_b$  denotes salinity at the freshwater, mixing zone and brackish sites, respectively. Evaporation and precipitation were assumed to be negligible (Ameryk *et al.*, 2005).

### *Chlorophyll-a, nutrients and bacterial abundance*

Biomass for measurements of chlorophyll-a concentration was collected by filtration of 100 - 200 ml of unfiltered water onto glass-fiber GF/F filters (average pore size 0.7 µm, Whatmann). The filters were stored in the dark at -20°C and analyzed within one month of collection. Chlorophyll-a was extracted for 24 h in 90% acetone in the dark at 4°C and measured using fluorometric methods (Evans *et al.*, 1987) with the Turner Designs 10-005R fluorometer. Concentrations of nutrients

( $\text{NO}_2^-$ ,  $\text{NO}_3^-$ ,  $\text{NH}_4^+$ ,  $\text{PO}_4^{3-}$ , dissolved Si, total P, total N) were determined from the unfiltered water samples with methods recommended for the Baltic Sea area (Grasshoff *et al.*, 1976).

Samples for bacterial abundance (10 ml) were fixed with 2% formalin for 1 h, and were stored at 4°C in the dark until further processing (< 30 days). Bacterial cells were stained with SYBR-Green I in TE buffer (10 mM Tris, 1 mM EDTA; Sigma) in the dark for 30 min and analyzed with an InFlux™ V-GS flow cytometer (BD, USA) with a blue laser (Coherent, Sapphire, 200 mW, 488 nm, detection wavelength 531 nm).

#### *Total RNA isolation and reverse transcription*

Two liters of the prefiltered water from each triplicate were filtered immediately after collection onto sterile polycarbonate filters (Whatman Cyclopore Track etched membrane, 47 mm diameter, 0.2 µm pore size, duration of the filtrations step ~10-30 minutes, depending on the sample) mounted in an autoclaved filtration tower. Filters were immediately frozen at -80°C and stored < 16 h. Total RNA was extracted with a GeneMATRIX Universal RNA Purification Kit (Eurx, Gdańsk, Poland) including a DNase digestion step (10 min at room temperature).

Reverse transcription was performed with dART reverse transcriptase (Eurx) using the Bact805R primer (Herlemann *et al.*, 2011) at 58°C for one hour and RNA was subsequently digested with RNase H (Eurx) for 30 min at 37°C.

#### *Amplification of 16S rRNA fragments and pyrosequencing*

Bacterial amplicons were prepared for pyrosequencing in a two-step process (Schülke, 2000). The V3-V4 16S rRNA fragments were amplified in the first PCR round from cDNA, with primers Bact341f and Bact805r (Herlemann *et al.*, 2011) bearing M13 and M13R overhangs, respectively (5 min at 95°C, 20 cycles of 40 s at 95°C, 40 s at 58°C, 1 min at 72°C, and final elongation of 7 min at 72°C) using a Pfu high-fidelity polymerase (Eurx). The PCR products were purified from agarose gels with GeneMATRIX Agarose-out DNA Purification Kit (Eurx). Barcodes and 454 pyrosequencing adapters were added in the subsequent PCR with M13 and M13R primers with overhangs bearing 10 bp barcode and adapter A and B, respectively (5 min at 94°C, 10 cycles of 30

s at 94°C, 45 s at 51°C, 30 s at 72°C; and final elongation of 10 min at 72°C). Barcodes differed by at least four nucleotides (Hamming distance = 4, Levenshtein distance = 4), which allows for correction of at least one error (in most cases two errors can be corrected) (Faircloth and Glenn, 2012).

The final products were gel-purified with Qiaquick Gel Extraction Kit (Qiagen) and their concentrations were measured using a PicoGreen kit (LifeTechnologies, Molecular Probes) on a Qbit fluorometer (LifeTechnologies). Samples were then pooled in equimolar amounts and pyrosequenced at the Centre for Genomic Research, University of Liverpool (Liverpool, UK) from both ends with the use of Lib-A emPCR kit and Titanium sequencing kit (Roche).

### *Bioinformatic analyses*

Bioinformatic analysis was performed with Mothur v.1.32 (Schloss *et al.*, 2009) and custom-tailored Perl scripts. The scripts are available from the authors upon request. The analyses were based on Schloss's SOP ([www.mothur.org/wiki/454\\_SOP](http://www.mothur.org/wiki/454_SOP)) with modifications necessary to process reads derived from both ends of the amplicons, increasing the effectiveness of denoising and chimera removal as well as producing 10 subsamples of the whole data and averaging the shared OTU table over those subsamples. For full description of the methodology see the Supplementary File 6.

Each OTU was assigned to one of the four categories based on the peak abundance of its reads (highest percentage of a given OTU): if the peaks consistently occurred in one zone in all seasons, the OTU was assigned to 'freshwater', 'mixing zone' or 'brackish' categories, accordingly. If the peak abundance was found in different zones of the estuary depending on the season, the OTU was assigned to the 'variable' category.

The core microbiome of a set of samples was defined as a set of OTUs that contributed at least 1% of reads to each sample within the set. It was identified with the `get.coremicrobiome` command of Mothur.

A relaxed neighbor joining (RNJ) tree was constructed from the final alignment with clearcut (clearcut, Sheneman *et al.*, 2006) and UniFrac (Lozupone and Knight, 2005) distance matrices were calculated in Mothur (unifrac.unweighted and unifrac.weighted) with subsampling of the RNJ tree to include 2500 reads per sample. Morisita-Horn (Horn, 1966) and Bray-Curtis (Bray and Curtis, 1957)  $\beta$  diversity distance matrices were calculated in R using vegan's function vegdist based on community matrix derived from Mothur-produced shared OTU table.

Furthermore, to check if OTUs delineated at the 97% similarity cutoff consisted of populations (subOTUs) that were restricted to particular zone of the estuary, annotated phylogenetic trees of higher phylogenetic resolution were constructed for the 50 most abundant OTUs. Reads belonging to the chosen OTUs were extracted from the whole set with Perl scripts, the distance matrix was calculated and OTUs were constructed as above, but using a 99% similarity cutoff. The annotated trees were constructed basing on most abundant with clearcut, as described above, and plotted with phyloseq's plot\_tree function.

The sequencing error rate was estimated basing on processing of the V4 fragment of 18S rRNA gene of *Skeletonema marinoi* BA98 amplified from genomic DNA isolated from pure culture with the seq.error command of mothur. The PR2 database (Guillou *et al.*, 2013) was used instead of SILVA for alignment and as the reference templates set for ChimeraSlayer. The *S. marinoi* BA98 18S rRNA gene sequence (HM805045.1) was used as a reference.

### *Statistical analyses*

Statistical analyses were performed in Mothur and R (R Development Core Team, 2011) using vegan (Oksanen *et al.*, 2013), phyloseq (McMurdie *et al.*, 2013), and Hmisc (Harell, 2014) packages.

Principal Component analysis of environmental variables was performed with the princomp function in R. Homogeneity of variance in environmental variables was assessed with PERMDISP test in vegan functions betadisper and anova, with 9999 permutations employed in the permutation test.

To check if communities differed significantly between the zones of the estuary, non-metric multidimensional scaling (nMDS) ordination based on UniFrac and Bray-Curtis distances combined with AMOVA (Excoffier *et al.*, 1992) and ANOSIM (Clarke, 1993) was performed, utilising 9999 permutations in both instances. To assess the influence of environmental variables on the community structure the envfit and cca functions of vegan were used. Here, the variables were square root transformed to decrease deviations from normality and 9999 permutations were used in the significance tests. Significance of differences in species richness, diversity and evenness was performed with ANOVA (aov function in R) with Tukey's HSD post-hoc analysis (TukeyHSD), assuming normality of distributions and homogeneity of variance.

### Acknowledgments

We thank Jerzy Lissowski, the cpt. of m/y Hestia, for help with sampling, and Lena Szymanek for preparation of the map of the Vistula estuary (Figure 7).

This study was supported by project 795/N-CBOL/2010/0 founded by the National Science Centre Poland, project DEMONA (PSPB-036/2010) supported by a grant from Switzerland through the Swiss contribution to the enlarged European Union (Polish-Swiss Research Programme), and project no. 15-12197S from Czech Science Foundation. Authors' contribution: MG – study design, bioinformatics, data analysis and interpretation of the results, writing the manuscript, JC – sampling, laboratory molecular work, SC – study design, bioinformatics, critical appraisal of the manuscript, KP – study design, sampling, interpretation of the results, writing the manuscript.

## References

- Ameryk, A., Hahnke, R.L., Gromisz, S., Kownacka, J., Zalewski, M., Szymanek, L., *et al.* (2014) Bacterial community structure influenced by *Coscinodiscus sp.* in the Vistula river plume. *Oceanologia* 56: 825-856.
- Ameryk, A., Podgorska, B., and Witek, Z. (2005) The dependence between bacterial production and environmental conditions in the Gulf of Gdańsk. *Oceanologia* 47: 27-45.
- Andersson, A.F., Riemann, L., and Bertilsson, S. (2010) Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *ISME J* 4: 171-181.
- Blazewicz S.J., Barnard R.L., Daly R.A., and Firestone M.K. (2013) Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *ISME J* 7: 2061-2068.
- Bray J.R., and Curtis J.T. (1957) An ordination of upland forest communities of southern Wisconsin. *Ecol Monographs* 27: 325-349.
- Campbell, B.J., and Kirchman, D.L. (2013) Bacterial diversity, community structure and potential growth rates along an estuarine salinity gradient. *ISME J* 7: 210-220.
- Chen H., Jogler M., Rohde M., Klenk H.P., Busse H.J., Tindall B.J., *et al.* (2013) *Sphingobium limneticum* sp. nov. and *Sphingobium boeckii* sp. nov., two freshwater planktonic member of the family Sphingomonadaceae, and reclassification of *Sphingomonas suberifaciens* as *Sphingobium suberifaciens* comb. nov. *Int J Syst Evol Microbiol* 63: 735-743.
- Clarke K.R. (1993) Non-parametric multivariate analyses of changes in community structure. *Aust J Ecol* 18: 117-143.
- Crump, B.C., Hopkinson, C.S., Sogin, M.L., and Hobbie, J.E. (2004) Microbial biogeography along an estuarine salinity gradient: Combined influences of bacterial growth and residence time. *Appl Environ Microbiol* 70:1494-1505.

Eren A.M., Maignien L., Sul W.J., Murphy L.G., Grim S.L., Morrison H.G, Sogin ML. (2013) Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol Evol* 4: doi: 10.1111/2041-210X.12114

Evans, C.A., O'Reilly, J.E., and Thomas J.P. (1987) A handbook for the measurement of chlorophyll-*a* and primary production. Biological investigations of marine antarctic systems and stocks (BIOMASSJ), Vol 8. Texas A&M University, College Station.

Excoffier L., Smouse P.E., and Quattro J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 470-491.

Faircloth B.C., and Glenn T.C. (2012) Not all sequence tag are created equal: designing and validating sequence identification tags robust to indels. *PLoS ONE* 7: e42543.

Grasshoff K., Kremling K., and Ehrhardt M. (1976) Methods of seawater analysis, 1st edn. Weinheim, Germany and New York, USA: Verlag Chemie.

Guillou L., Bachar D., Audic S., Bass D., Berney C., Bittner L., *et al.* (2013) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small subunit rRNA sequences with curated taxonomy. *Nucl Acids Res* 41: D597-D604.

Harrell F.E. (2014) Hmisc: Harrell Miscellaneous. R package version 3.14-5. <http://CRAN.R-project.org/package=Hmisc>.

Herlemann, D.P.R., Labrenz, M., Juergens, K., Bertilsson, S., Waniek, J.J., and Andersson, A.F. (2011) Transitions in bacterial communities along the 2000 km salinity gradient of the baltic sea. *ISME J.* 5: 1571-1579.

Horn H. (1966) Measuring of “overlap” in comparative ecological studies. *Am Nat* 100: 419-424.

Ibarbalz F.M., Perez M.V., Figuerola E.L.M., and Erijman L. (2014) Bias associated with amplicon sequencing does not affect the quantitative assessment of bacterial community dynamics. *PLoS ONE* 9: e99722.

Kowalkowski, T., Pastuszak, M., Igras, J., and Buszewski, B. (2012) Differences in emission of nitrogen and phosphorus into the Vistula and Oder basins in 1995-2008-natural and anthropogenic causes (moneris model). *J Marine Systems* 89: 48-60.

Li C., Lai Q., Li G., Liu Y., Sun F., and Shao Z. (2014) Multilocus sequence analysis for the assessment of phylogenetic diversity and biogeography in *Hyphomonas* bacteria from diverse marine environments. *PLoS ONE* 9: e101394.

Lindh M.V., Figueroa D., Sjoestedt J., Baltar F., Lundin D., Andersson A., Legrand C, Pinhassi J (2015) Transplant experiments uncover Baltic Sea basin-specific responses in bacterioplankton community composition and metabolic activities. *Front Microbiol* doi: 10.3389/fmicb.2015.00223

Logares, R., Bråte, J., Bertilsson, S., Clasen, J. L., Shalchian-Tabrizi, K., and Rengefors, K. (2009) Infrequent marine-freshwater transitions in the microbial world. *Trends Microbiol* 17: 414-422.

Logares, R., Bråte, J., Heinrich F., Shalchian-Tabrizi K., and Bertilsson S. (2010) Infrequent transitions between saline and fresh waters in one of the most abundant microbial lineages (SAR11). *Mol Biol Evol* 27: 347-357.

Lozupone C., and Knight R. (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71: 8228-8235.

McMurdie P.J., and Holmes S. (2013) phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* 8: e61217.

Newton R.J., Jones S.E., Eller A., McMahon K.D., and Bertilsson S. (2011) A guide to the natural history of freshwater lake bacteria. *Microbiol Mol Biol Rev* 75: 14-49.

Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'Hara R.B., *et al.* (2013) vegan: Community Ecology Package. R package version 2.0-10. URL <http://CRAN.R-project.org/package=vegan>.

Park S., Park J.M., Kang C.H., and Yoon J.H. (2014) *Confluentimicrobium lipolyticum* gen. nov., sp. nov., a novel lipolytic alphaproteobacterium isolated from the junction between the ocean



- and freshwater spring, and emended description of *Actibacterium Mucosum* Lucena *et al.* 2012. *Antonie Van Leeuwenhoek* 106: 969-977.
- Piwoż, K., Salcher, M.M., Zeder, M., Ameryk, A., and Pernthaler, J. (2013) Seasonal dynamics and activity of typical freshwater bacteria in brackish waters of the gulf of gdańsk. *Limnol Oceanogr* 58: 817-826.
- R Development Core Team (2011) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Remane A. (1934) Die Brackwasserfauna. *Zool Anz Suppl* 7: 34-74.
- Riemann, L., Leitet, C., Pommier, T., Simu, K., Holmfeldt, K., Larsson, U., and Hagström, A. (2008) The native bacterioplankton community in the central Baltic Sea is influenced by freshwater bacterial species. *Appl Environ Microbiol* 74: 503-515.
- Ruiz A., Franco J., and Villate F. (1998) Microzooplankton grazing in the estuary of Mundaka, Spain, and its impact on phytoplankton distribution along the salinity gradient. *Aquat Microb Ecol*, 14: 281-288.
- Ryall B., Eydallin G., and Ferenci T. (2012) Culture history and population heterogeneity as determinants of bacterial adaptation: the adaptomics of a single environmental transition. *Microbiol Mol Biol Rev* 76: 597-625.
- Ryan F.J. (1952) Adaptation to use lactose in *Escherichia coli*. *J Gen Microbiol* 7: 69-88.
- Schloss P.D., Westcott S.L., Ryabin T., Hall J.R., Hartmann M., Hollister E.B., *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75: 7537-7541.
- Sheneman L., Evans J., and Foster J.A. (2006) Clearcut: a fast implementation of relaxed neighbor joining. *Bioinformatics* 22: 2823-2824.

Telesh I.V., Schubert H., and Skarlato S.O. (2011a) Revisiting Remane's concept: evidence for high plankton diversity and a protistan species maximum in the horohalinicum of the Baltic Sea. *Marine Ecol Prog Ser* 421: 1-11.

Telesh I.V., Schubert H., and Skarlato S.O. (2011b) Protistan diversity does peak in the horohalinicum of the Baltic Sea: Reply to Ptacnik *et al.* (2011) *Marine Ecol Prog Ser* 432: 293-297.

Theroux S., Huang Y., and Amaral-Zettler L. (2012) Comparative molecular microbial ecology of the spring haptophyte bloom in a Greenland arctic oligosaline lake. *Frontiers Microb* 3: Article 425.

Turbaugh P.J., Hamady M., Yatsunenko T., Cantarel B.L., Duncan A., Ley R.E., *et al.* (2009) A core gut microbiome in obese and lean twins. *Nature* 457: 480-484.

Wielgat-Rychert, M., Ameryk, A., Jarosiewicz, A., Kownacka, J., Rychert, K., Szymanek, L., *et al.* (2013) Impact of the inflow of Vistula river waters on the pelagic zone in the Gulf of Gdańsk. *Oceanologia* 55: 859-886.

Witek, Z., Humborg, C., Savchuk, O., Grelowski, A., and Łysiak-Pastuszak, E. (2003) Nitrogen and phosphorus budgets of the Gulf of Gdańsk (Baltic Sea). *Estuar Coast Shelf Sci* 57: 239-248.

Zarraonaindia I., Owens S.M., Weisenhorn P., West K., Hampton-Marcell J., Lax S., *et al.* (2015) The soil microbiome influences grapevine-associated microbiota. *mBio* 6(2): e02527-14.

Zhang G., Yang Y., Wang S., Sun Z., and Jiao K. (2015) *Alkalimicrobium pacificum* gen. nov., sp. nov., a marine bacterium in the family Rhodobacteraceae. *Int J Syst Evol Microbiol* 65: 2453-2458.

Zhao Z., Sun J., Zhang R., and Jiao N (2016) *Binariimonas pacifica* gen. nov., sp. nov., a novel marine bacterium of family Sphingomonadaceae isolated from east Pacific Ocean surface seawater. *Curr Microbiol* 72: 276-281.

## Tables

**Table 1.** Environmental and biological parameters at the sampling sites. Average  $\pm$  standard deviation values based on triplicate samples are given. All values were rounded to one decimal place, and value of standard deviation = 0.0 indicates that it was  $< 0.05$ . T- temperature in  $^{\circ}\text{C}$ , S – salinity, N-tot – total nitrogen in  $\mu\text{M}$ , DIN – dissolved inorganic nitrogen in  $\mu\text{M}$ , P-tot – total phosphorus in  $\mu\text{M}$ , SRP – soluble reactive phosphorus in  $\mu\text{M}$ , DSi – dissolved silica in  $\mu\text{M}$ , Chl-a – total chlorophyll-*a* in  $\mu\text{g l}^{-1}$ . Bacterial abundance (Bact. abund.) is given in  $10^6$  cells  $\text{ml}^{-1}$ . F – freshwater station in the Vistula River, MZ – mixing zone station, B – brackish station in the Gulf of Gdańsk

Date	site	T	S	N-tot	DIN	P-tot	SRP	DSi	Chl-a	Bact. abund.
05 Jul	F	19.0	0.5	37.6 $\pm$ 11.6	2.3 $\pm$ 0.8	1.2 $\pm$ 0.0	0.6 $\pm$ 0.1	63.5 $\pm$ 0.5	77.0 $\pm$ 3.7	5.3 $\pm$ 0.4
07 Jul	MZ	18.3	3.0	1.3 $\pm$ 0.1	0.9 $\pm$ 0.1	0.3 $\pm$ 0.0	0.1 $\pm$ 0.0	25.5 $\pm$ 3.4	49.8 $\pm$ 27.7	5.1 $\pm$ 2.0
07 Jul	B	16.6	7.1	15.4 $\pm$ 0.6	1.2 $\pm$ 0.2	0.1 $\pm$ 0.1	0.1 $\pm$ 0.0	5.5 $\pm$ 1.0	3.8 $\pm$ 0.6	1.1 $\pm$ 0.2
17 Oct	F	9.7	0.4	85.3 $\pm$ 2.6	44.3 $\pm$ 1.2	3.5 $\pm$ 0.1	2.1 $\pm$ 0.2	127.5 $\pm$ 5.0	8.8 $\pm$ 0.8	3.0 $\pm$ 0.9
19 Oct	MZ	10.8	3.2	56.1 $\pm$ 1.1	22.8 $\pm$ 0.1	2.2 $\pm$ 0.0	1.5 $\pm$ 0.0	77.9 $\pm$ 1.4	5.9 $\pm$ 0.1	1.4 $\pm$ 0.4
19 Oct	B	13.0	6.9	22.4 $\pm$ 3.3	2.2 $\pm$ 0.4	0.7 $\pm$ 0.0	0.4 $\pm$ 0.0	10.2 $\pm$ 0.4	1.7 $\pm$ 0.1	1.2 $\pm$ 0.2
23 Jan	F	1.8	0.5	165.5 $\pm$ 2.3	94.9 $\pm$ 4.3	3.5 $\pm$ 0.0	2.1 $\pm$ 0.1	189.1 $\pm$ 6.2	2.6 $\pm$ 0.1	0.9 $\pm$ 0.1
25 Jan	MZ	1.6	2.6	126.1 $\pm$ 47.2	75.7 $\pm$ 29.1	2.8 $\pm$ 0.6	2.0 $\pm$ 0.2	164.3 $\pm$ 50.6	2.8 $\pm$ 0.3	0.8 $\pm$ 0.2
25 Jan	B	3.6	7.6	22.4 $\pm$ 6.9	4.2 $\pm$ 1.3	0.6 $\pm$ 0.0	0.6 $\pm$ 0.0	8.5 $\pm$ 1.0	0.9 $\pm$ 0.0	0.4 $\pm$ 0.1
13 Apr	F	8.5	0.3	76.8 $\pm$ 1.2	62.2 $\pm$ 3.5	3.2 $\pm$ 0.0	1.0 $\pm$ 0.1	89.1 $\pm$ 2.9	42.5 $\pm$ 1.8	4.3 $\pm$ 1.0
17 Apr	MZ	6.9	4.1	42.7 $\pm$ 2.9	24.8 $\pm$ 2.4	2.3 $\pm$ 0.2	0.8 $\pm$ 0.1	54.8 $\pm$ 10.2	31.9 $\pm$ 1.1	1.4 $\pm$ 0.1
17 Apr	B	6.2	7.5	9.7 $\pm$ 1.5	0.6 $\pm$ 0.2	0.6 $\pm$ 0.0	0.2 $\pm$ 0.1	3.2 $\pm$ 0.5	2.2 $\pm$ 0.7	0.3 $\pm$ 0.1

**Table 2.** Deviations of the theoretical values of environmental variables from the measured values (in percent relative to the measured values). These theoretical concentrations were computed from fractions of fresh water in the mixing zone that were calculated from salinity. Vistula waters: fraction of the freshwaters in the mixing zone, Bacteria – bacterial abundance, P-tot – total phosphorus, N-tot – total nitrogen, DSi – dissolved silica in  $\mu\text{M}$ , Chl-a –chlorophyll-*a*.

Date	Vistula waters	Bacteria	P <sub>tot</sub>	N <sub>tot</sub>	DSi	Chl- <i>a</i>
17 Apr 2012	0.6375	27.45	3.36	-22.98	-5,86	12.57
07 Jul 2011	0.6197	29.59	-175.75	-2178.29	-62,73	1.28
19 Oct 2011	0.5626	7.29	-3.75	-2.91	2.15	3.48
25 Jan 2012	0.7022	13.54	5.58	2.53	17.62	25.22

**Table 3.** Summer mixing zone OTUs with number of reads higher than expected from the passive mixing alone. Significance tested with ANOVA, with homogeneity of variance and normality assumed.  $P < 0.05$  in all cases.

OTU	Factor <sup>a</sup>	Taxonomy	Tribe <sup>b</sup>
OTU3	5.9	unclassified member of Sphingomonadaceae	-
OTU16	11.2	<i>Brevundimonas</i>	Brev
OTU19	6.4	<i>Hyphomonas</i>	-
OTU24	2.6	<i>Roseibacterium</i>	-
OTU27	2.4	unclassified member of MNG7 (Rhizobiales)	-
OTU36	6.5	<i>Caulobacter</i>	Brev
OTU38	1.7	<i>Hirschia</i>	-
OTU42	17.3	<i>Arcobacter</i>	-
OTU60	3.6	unclassified member of Flavobacteriaceae	-
OTU68	2.5	<i>Zymomonas</i>	-
OTU73	5.4	<i>Novosphingobium</i>	Novo-A2
OTU92	7.0	<i>Phenylobacterium</i>	Brev
OTU97	6.2	<i>Rhodobacter</i>	-

a – Factor indicating how many times the number of reads of a given OTUs was higher than expected from the passive mixing alone.

b – Tribe membership according to Newton *et al.* (2011).

**Table 4.** Freshwater OTUs detected at the brackish site in Gulf of Gdansk. OTUs with overall abundance in a given season >1% were listed.

Season	OTUs	Taxonomy	Tribe <sup>a</sup>
Spring	OTU1	Actinobacteria;Frankiales;Sporichthyaceae;hgcI_clade	-
	OTU3	Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;	-
	OTU4	Betaproteobacteria;Burkholderiales;Comamonadaceae; <i>Limnohabitans</i>	-
	OTU5	Actinobacteria;Frankiales;Sporichthyaceae;	-
	OTU6	Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;	-
	OTU9	Betaproteobacteria;Burkholderiales;Comamonadaceae; <i>Polaromonas</i>	-
	OTU16	Alphaproteobacteria;Caulobacterales;Caulobacteraceae; <i>Brevundimonas</i>	Brev
	OTU20	Betaproteobacteria;Burkholderiales;Alcaligenaceae;GKS98_fres hwater_group	betIII-A1
	OTU40	Actinobacteria;Frankiales;Sporichthyaceae;hgcI_clade	acI-A4
	OTU45	Flavobacteriia;Flavobacteriales;Flavobacteriaceae; <i>Flavobacterium</i>	Flavo-A2
Summer	OTU50	Actinobacteria;Frankiales;Sporichthyaceae;hgcI_clade	acI-A5
	OTU66	Flavobacteriia;Flavobacteriales;Flavobacteriaceae; <i>Flavobacterium</i>	-
	OTU1	Actinobacteria;Frankiales;Sporichthyaceae;hgcI_clade	-
Autumn	OTU3	Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;	-
	OTU5	Actinobacteria;Frankiales;Sporichthyaceae;	-
	OTU13	Acidimicrobiia;Acidimicrobiales;Acidimicrobiaceae;CL500- 29_marine_group	Iluma-A2

	OTU25	Actinobacteria;Frankiales;Sporichthyaceae;hgcI_clade	acI-B1
	OTU1	Actinobacteria;Frankiales;Sporichthyaceae;hgcI_clade	
	OTU3	Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;	-
	OTU4	Betaproteobacteria;Burkholderiales;Comamonadaceae;	-
		<i>Limnohabitans</i>	
Winter	OTU5	Actinobacteria;Frankiales;Sporichthyaceae;	-
	OTU20	Betaproteobacteria;Burkholderiales;Alcaligenaceae;GKS98	betIII-A1
		freshwater group	
	OTU42	Epsilonproteobacteria;Campylobacterales;Campylobacteraceae;	-
		<i>Arcobacter</i>	

---

a – Tribe membership according to Newton et al. (2011).

## Figure captions

Figure 1: Principal component analysis of environmental variables. Shape denotes site, and shade season.

Figure 2: Bacterial diversity and species richness in Vistula river estuary. A – observed number of OTUs, B – estimated total number of OTUs (Chao1 index), C – Shannon's H', D – Shannon's evenness.

Figure 3: Taxonomic composition of bacterial communities. A – phylum level, B – class level, C – family level, D – genus level. F – freshwater samples, MZ – mixing zone samples, B – brackish samples. Rare – sum of numbers of reads in taxa with abundance below 1%. The percentage of unclassified reads is lower at the genus level than at the family level due to some genera being affiliated to a higher taxonomic rank instead to a family. Average values from triplicates are given.

Figure 4: Ordination of communities from Vistula estuary. A – NMDS on UniFrac distance matrix, B – CCA on Bray-Curtis distance matrix. Triangles denotes freshwater, diamonds mixing zone and pentagons brackish, fill colour denotes season. Circles represent OTUs, size of which is proportional to the percentage contribution of reads in the total dataset, the fill colour denotes phylum and the border colour OTU type assigned based on its peak abundance. Numerical identifiers (1 for OTU1, etc.) of the most abundant OTUs characteristic for different zones of the estuary are shown next to the circles representing them. The following clusters of samples were significantly separated according to AMOVA and ANOSIM: B-F, B-MZ ( $p < 1.7 \times 10^{-2}$ , Bonferroni corrected, shown on panel A), spring-summer, spring-autumn, summer-winter, autumn-winter ( $p < 8.3 \times 10^{-3}$ , Bonferroni corrected, shown on panel B).



Figure 5: Contribution of freshwater, mixing zone and brackish phylotypes to the mixing zone and brackish communities. A – percentage of OTUs from different groups in the mixing zone, B – percentage of reads coming from different groups of OTUs in the mixing zone, C – percentage of OTUs from different groups in the Gulf of Gdansk, D – percentage of reads coming from different groups of OTUs in the Gulf of Gdansk. Freshwater share denotes share of water masses from the Vistula River in the mixing zone.

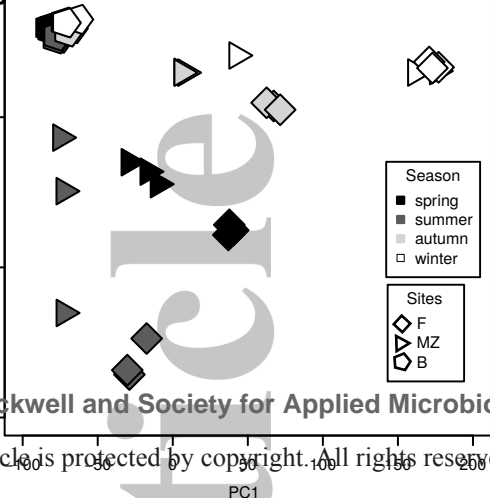
Figure 6: Annotated Relaxed Neighbor Joining trees for subOTUs of OTU20 (A) and OTU18 (B). The barcharts represent abundance of a given subOTU in different zones of the estuary.

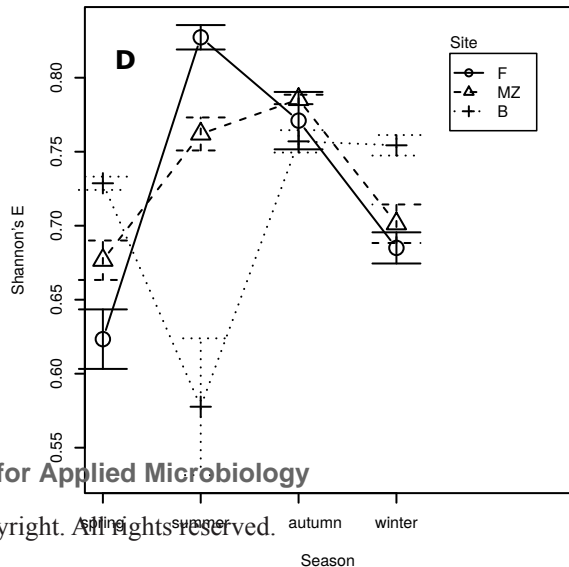
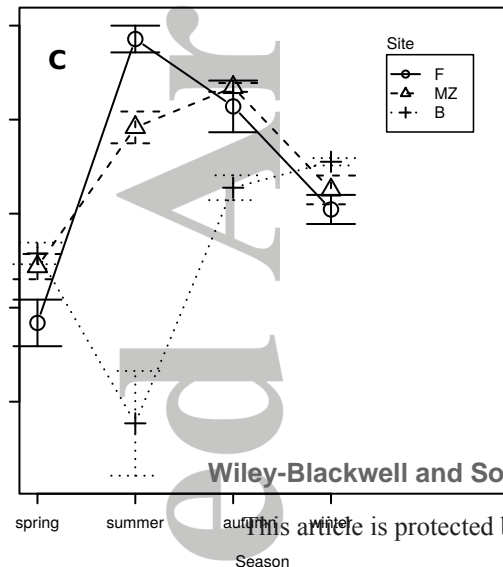
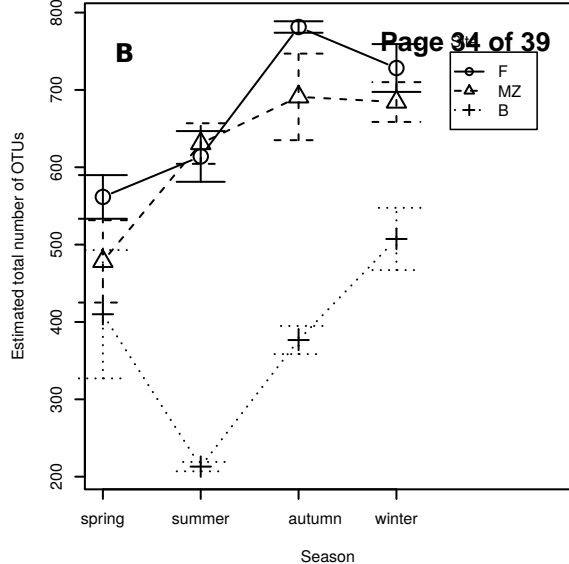
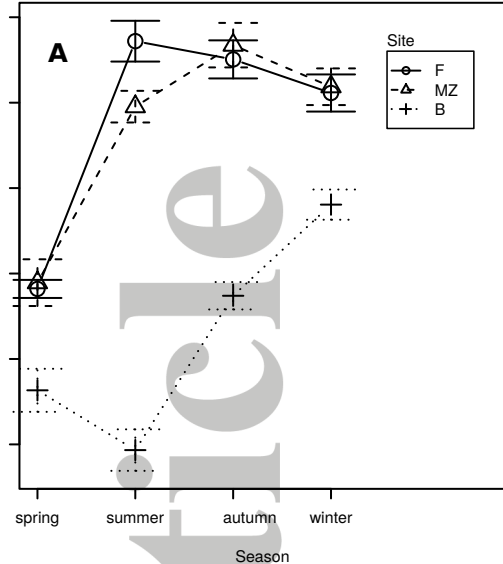
Figure 7: Location of the sampling stations. A – Map of the Baltic Sea with the Gulf of Gdańsk marked by the rectangle; B – the sampling stations: F – freshwater station at Kiezmark, MZ – mixing zone station at the mouth of Vistula, B – brackish station at the Gulf of Gdańsk.

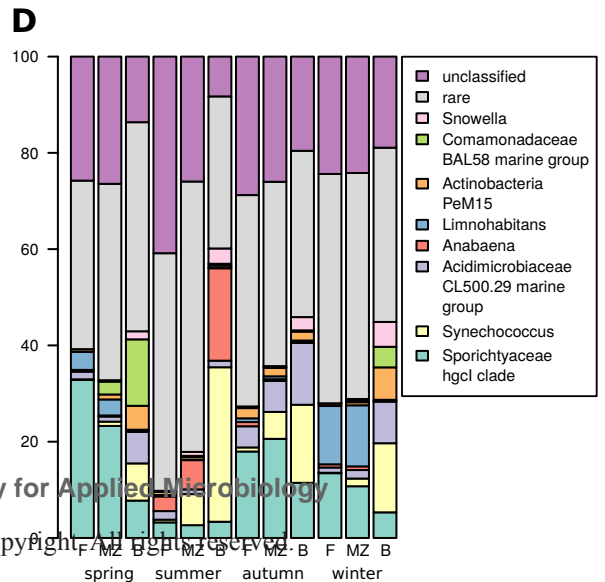
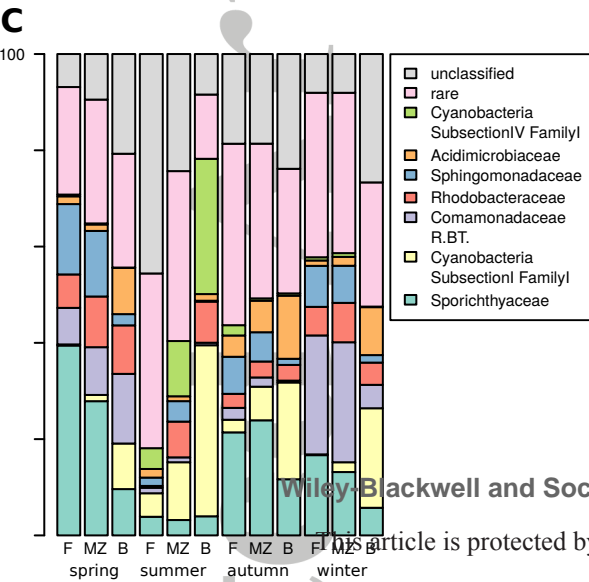
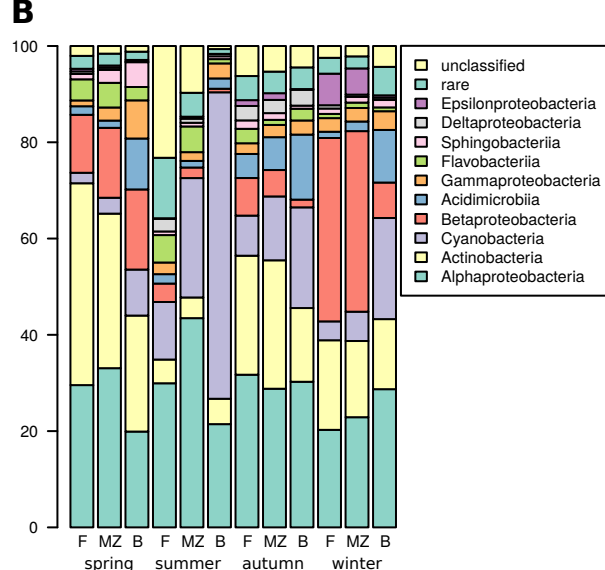
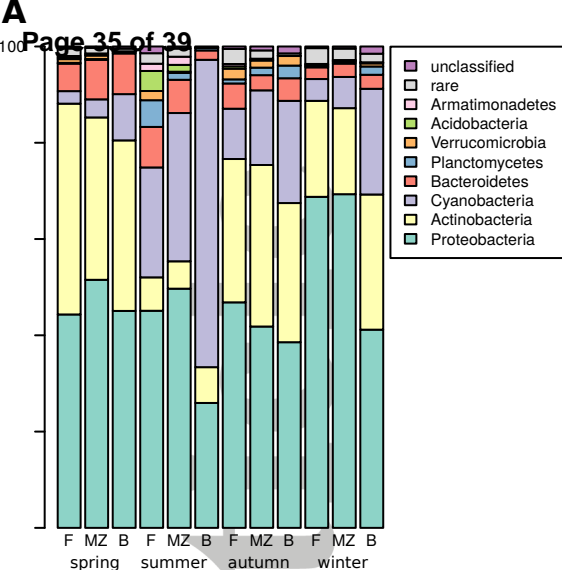
PC2

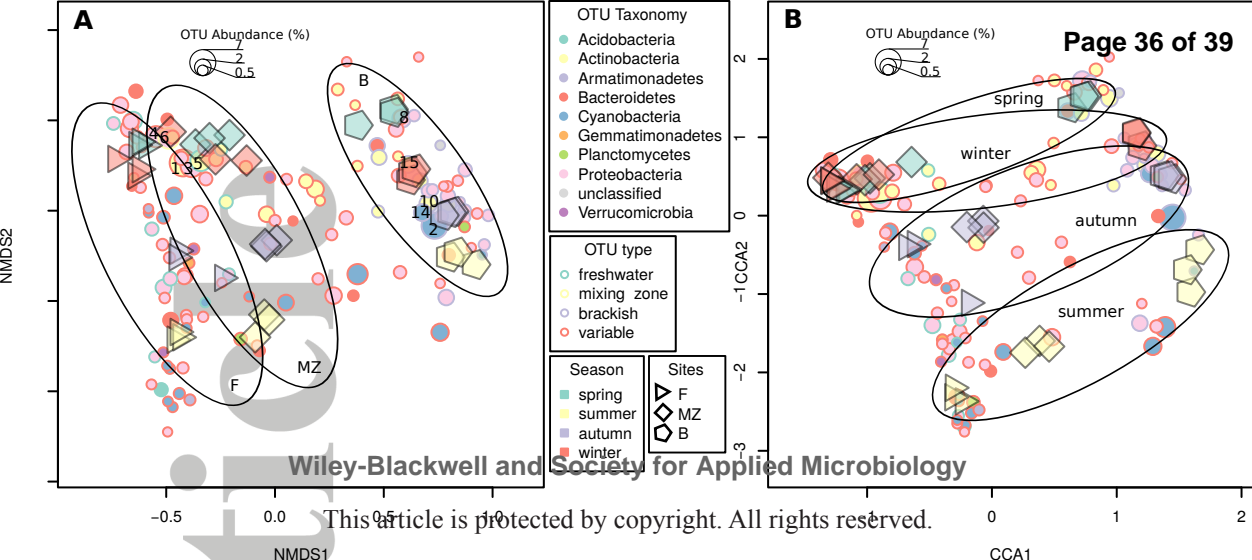
y-Blackwell and Society for Applied Microbiology

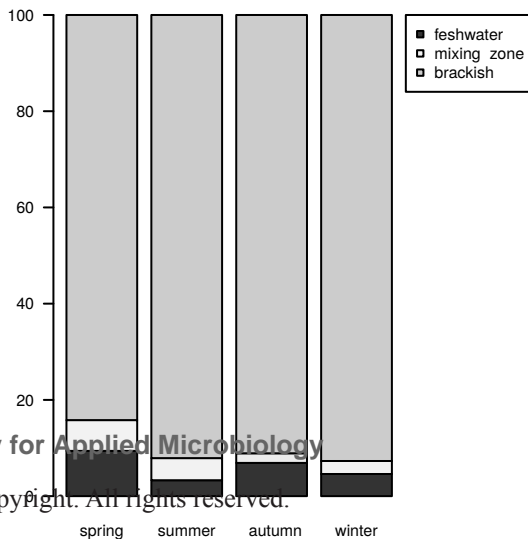
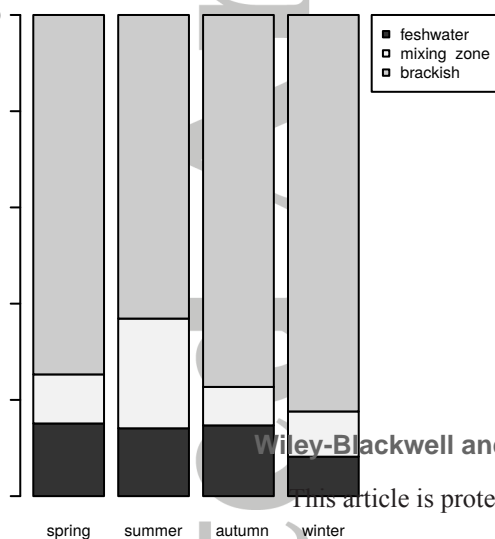
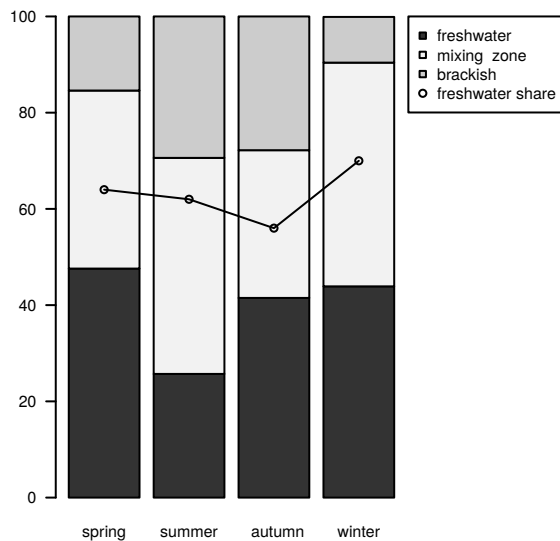
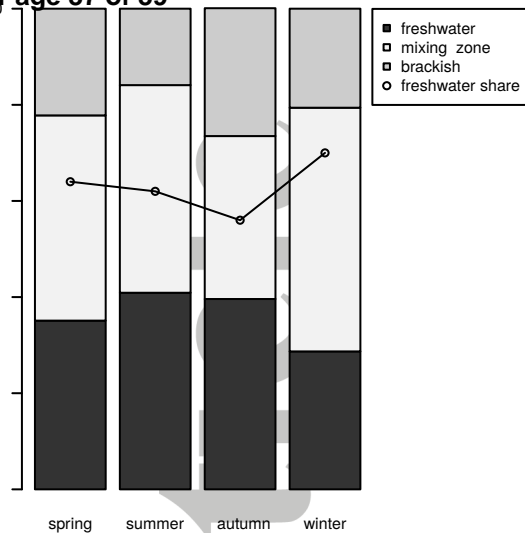
This article is protected by copyright. All rights reserved.

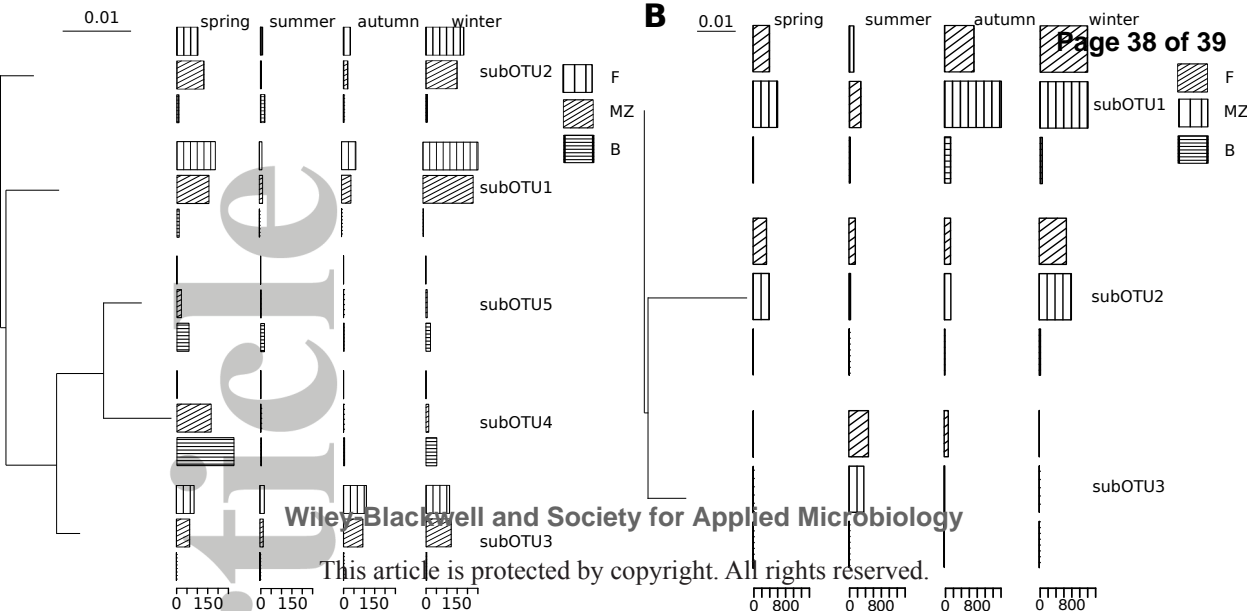




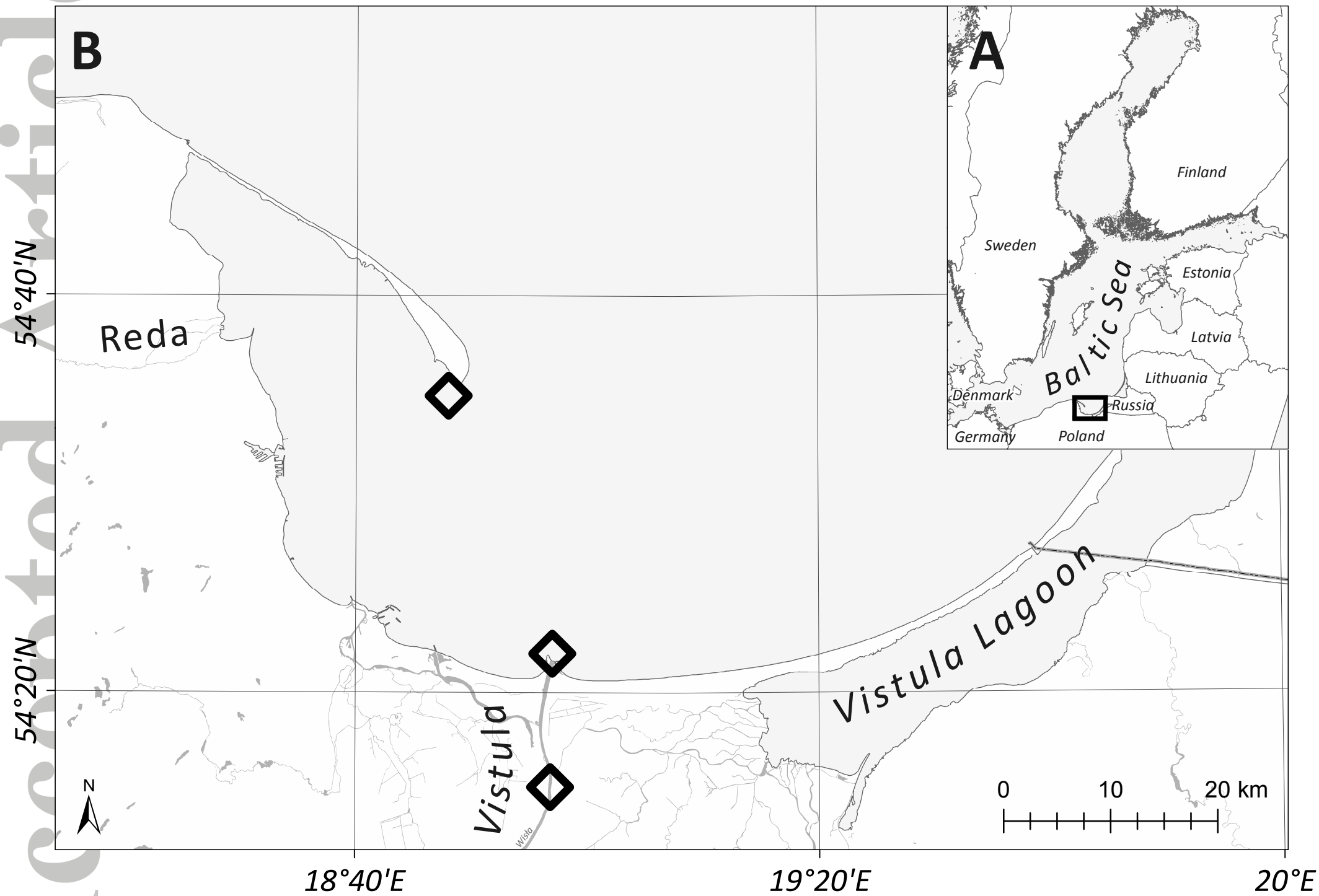




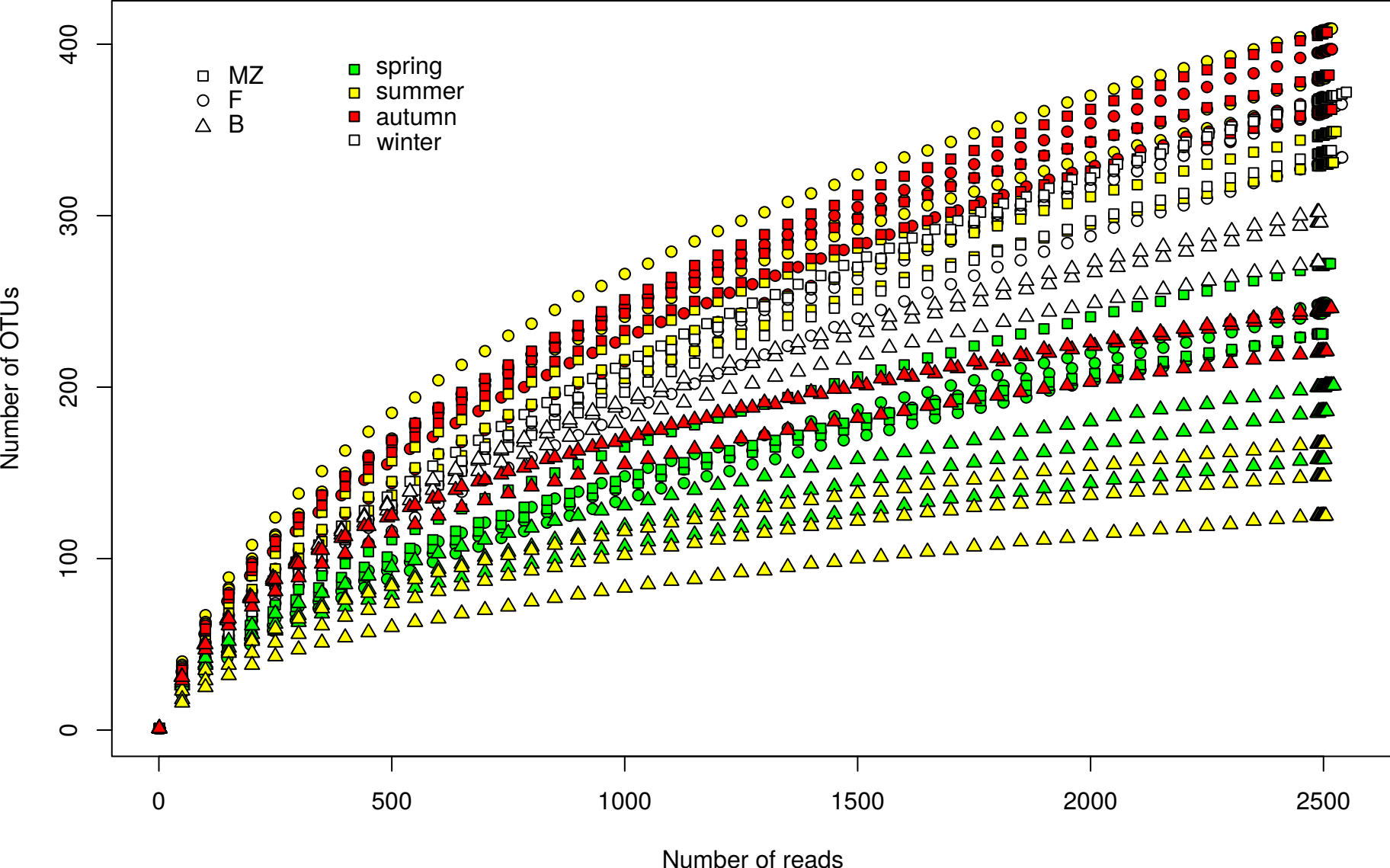




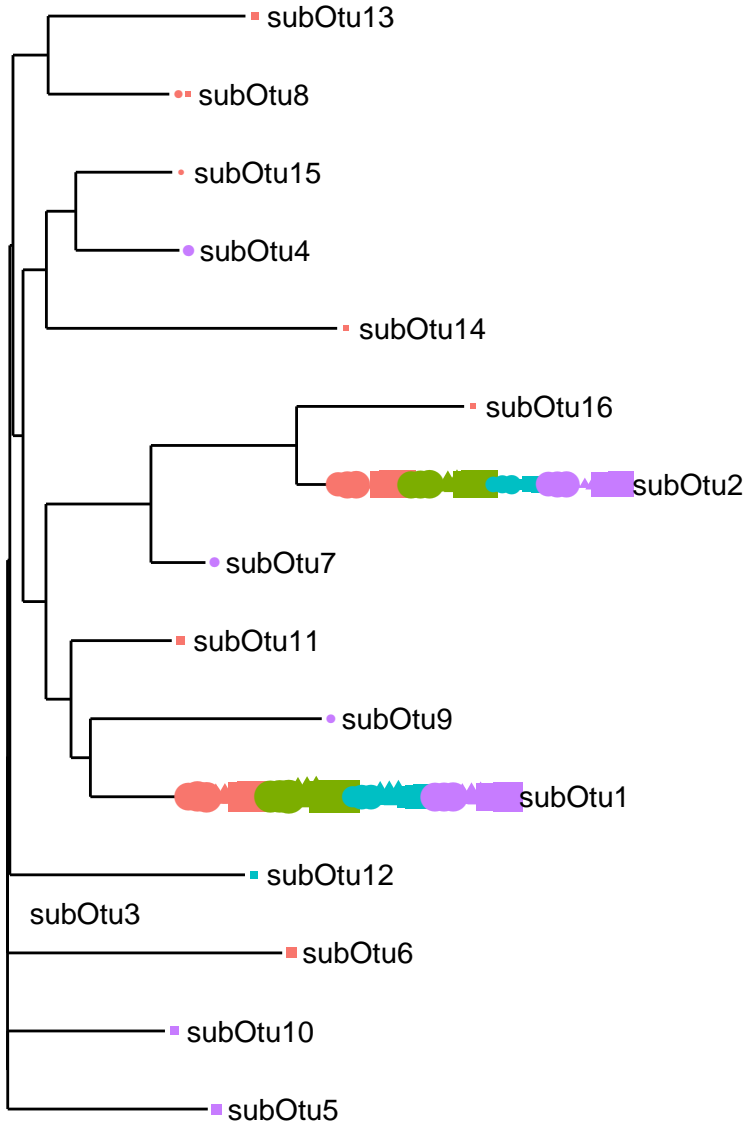
Accepted Article







otu1



**site**

- freshwater
- ▲ marine
- mixing\_zone

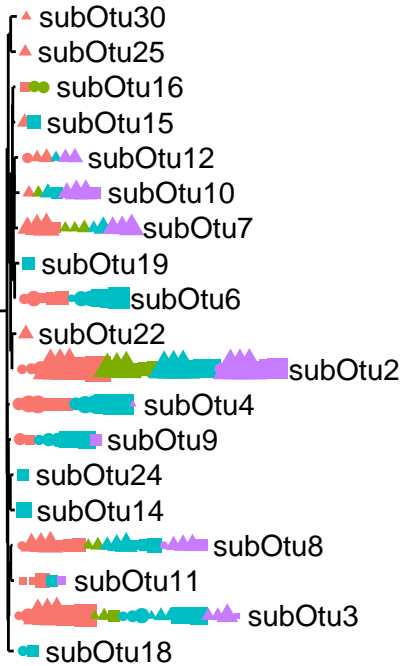
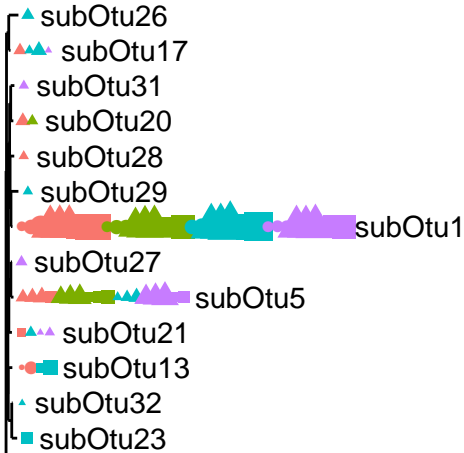
**season**

- fall
- spring
- summer
- winter

**Abundance**

- 25
- 625

# otu2



## site

- freshwater
- ▲ marine
- mixing\_zone

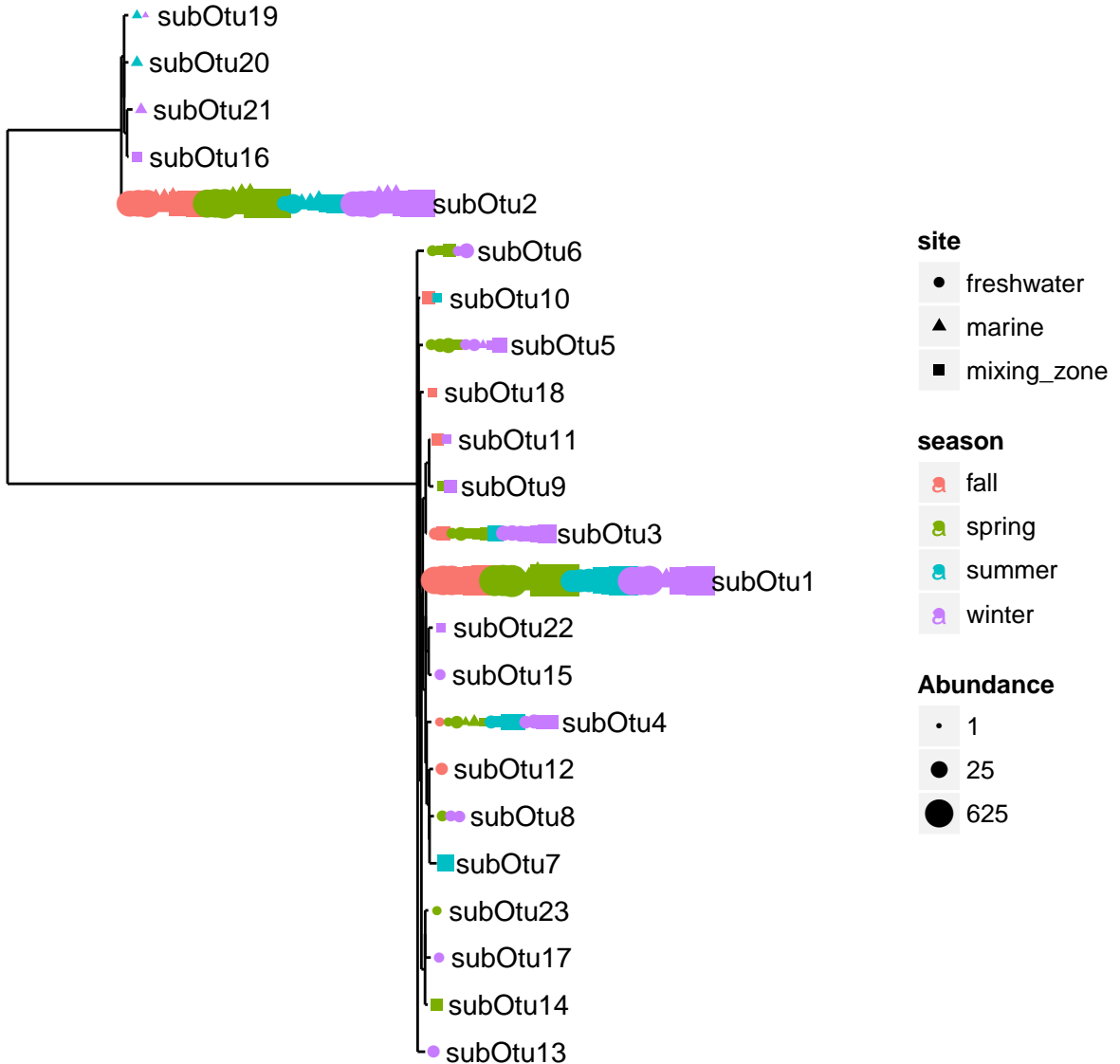
## season

- fall
- spring
- summer
- winter

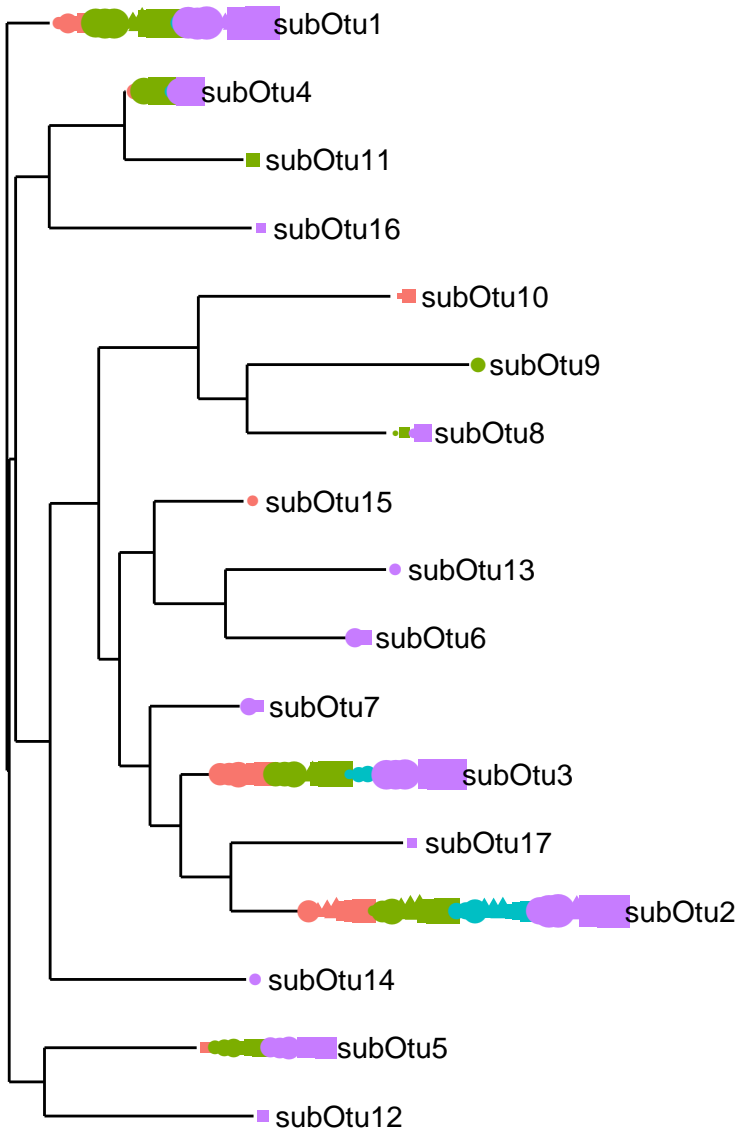
## Abundance

- 1
- 25
- 625

# otu3



# otu4



## site

- freshwater
- ▲ marine
- mixing\_zone

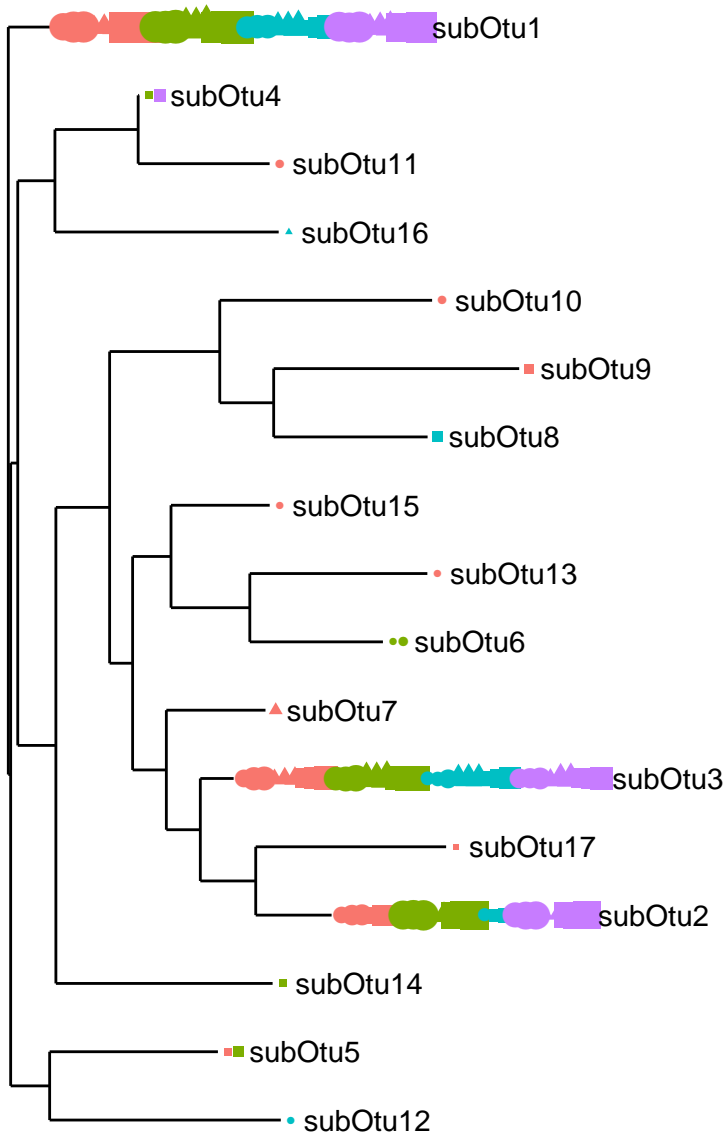
## season

- fall
- spring
- summer
- winter

## Abundance

- 1
- 25
- 625

# otu5



## site

- freshwater
- ▲ marine
- mixing\_zone

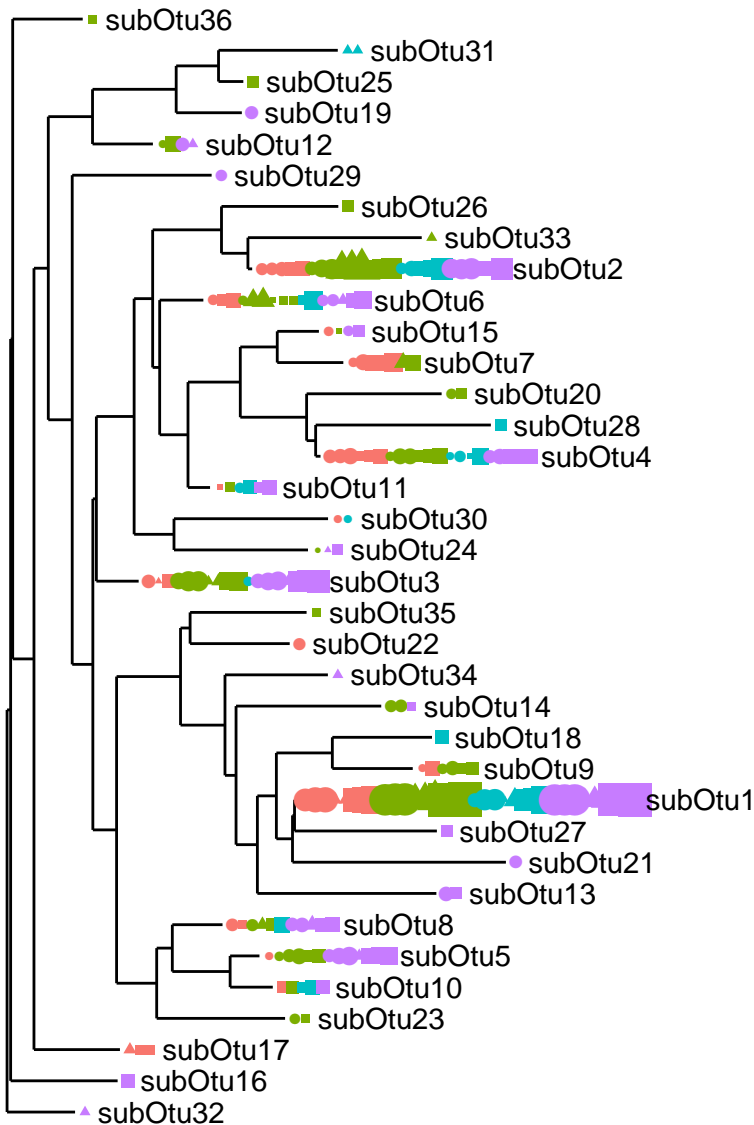
## season

- fall
- spring
- summer
- winter

## Abundance

- 25
- 625

# otu6



## site

- freshwater
- ▲ marine
- mixing\_zone

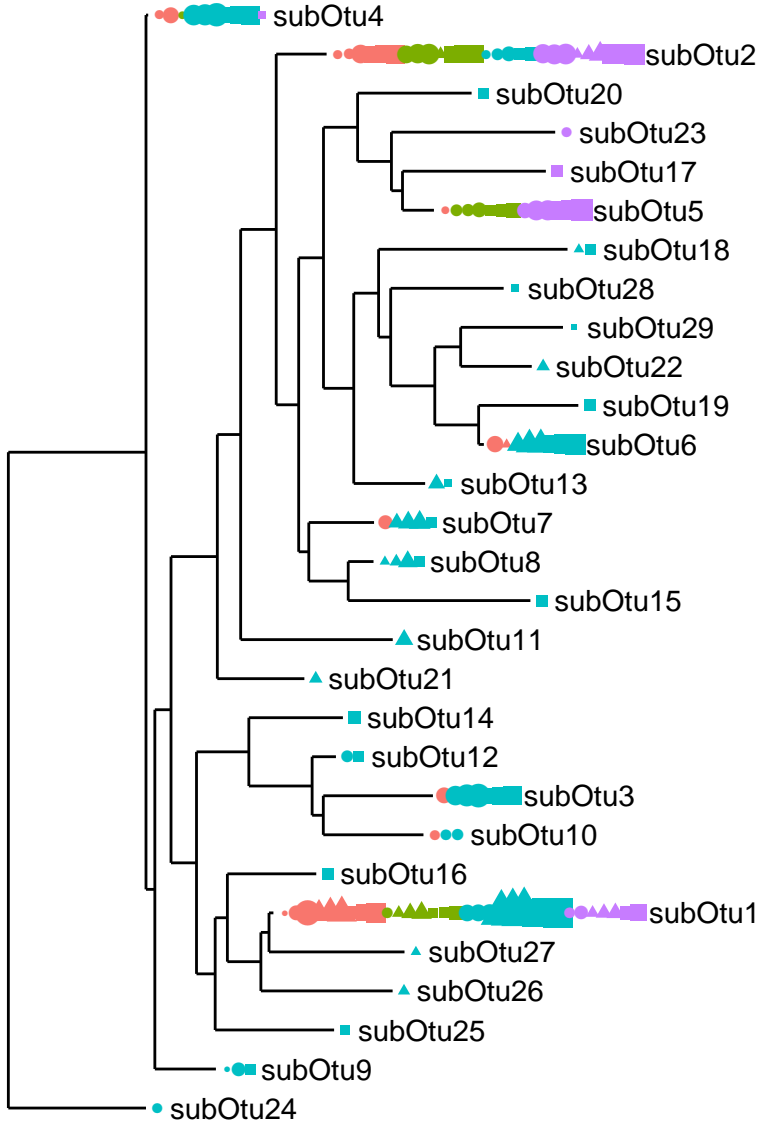
## season

- fall
- spring
- summer
- winter

## Abundance

- 1
- 25
- 625

# otu7



## site

- freshwater
- ▲ marine
- mixing\_zone

## season

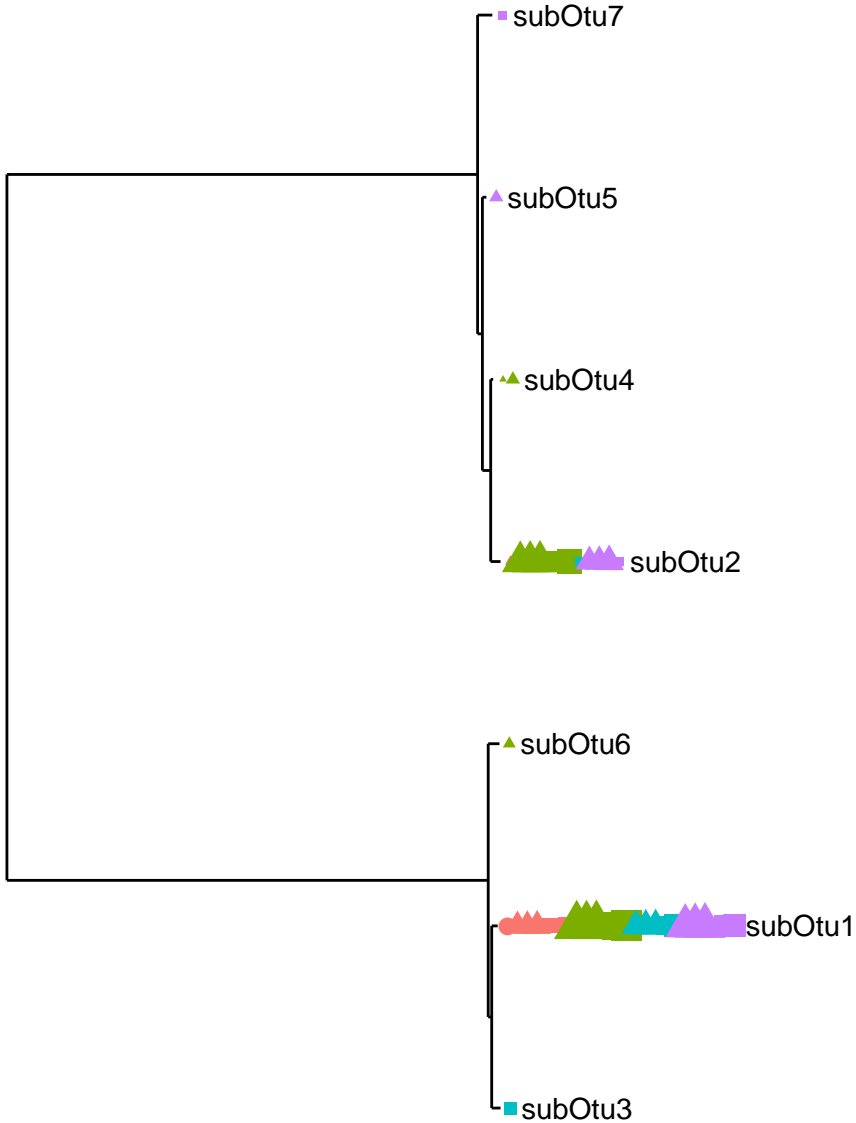
- fall
- spring
- summer
- winter

## Abundance

- 1
- 25
- 625



otu8



**site**

- freshwater
- ▲ marine
- mixing\_zone

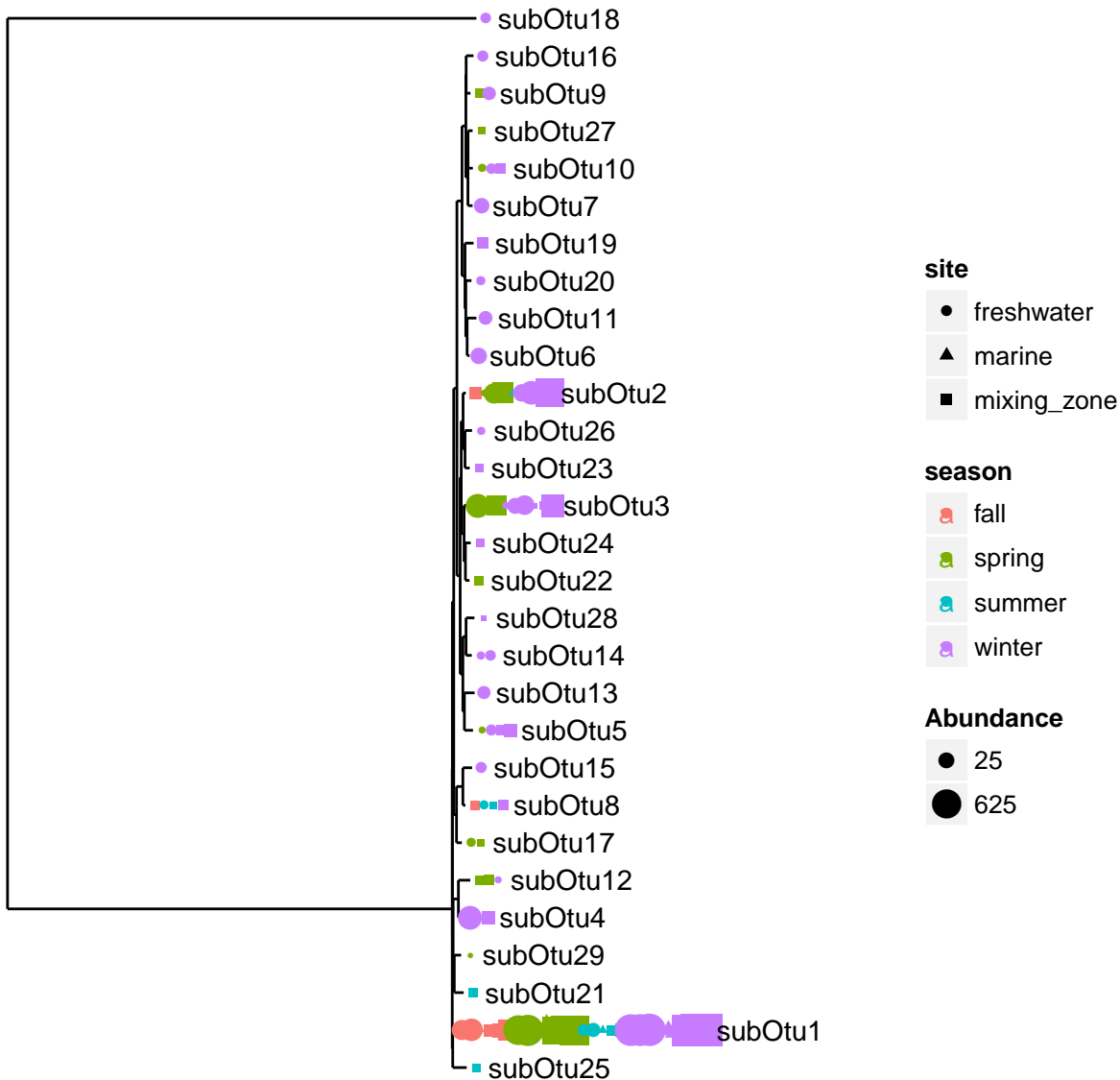
**season**

- fall
- spring
- summer
- winter

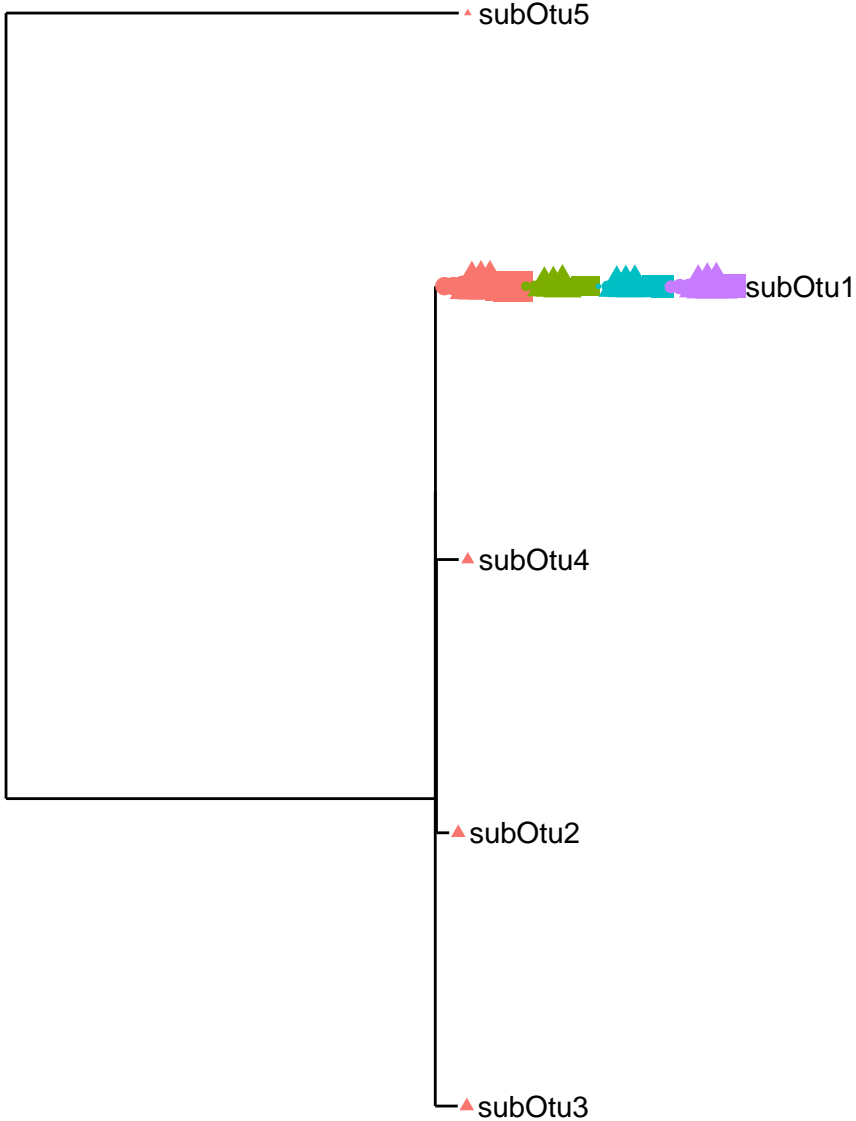
**Abundance**

- 1
- 25
- 625

otu9



otu10



**site**

- freshwater
- ▲ marine
- mixing\_zone

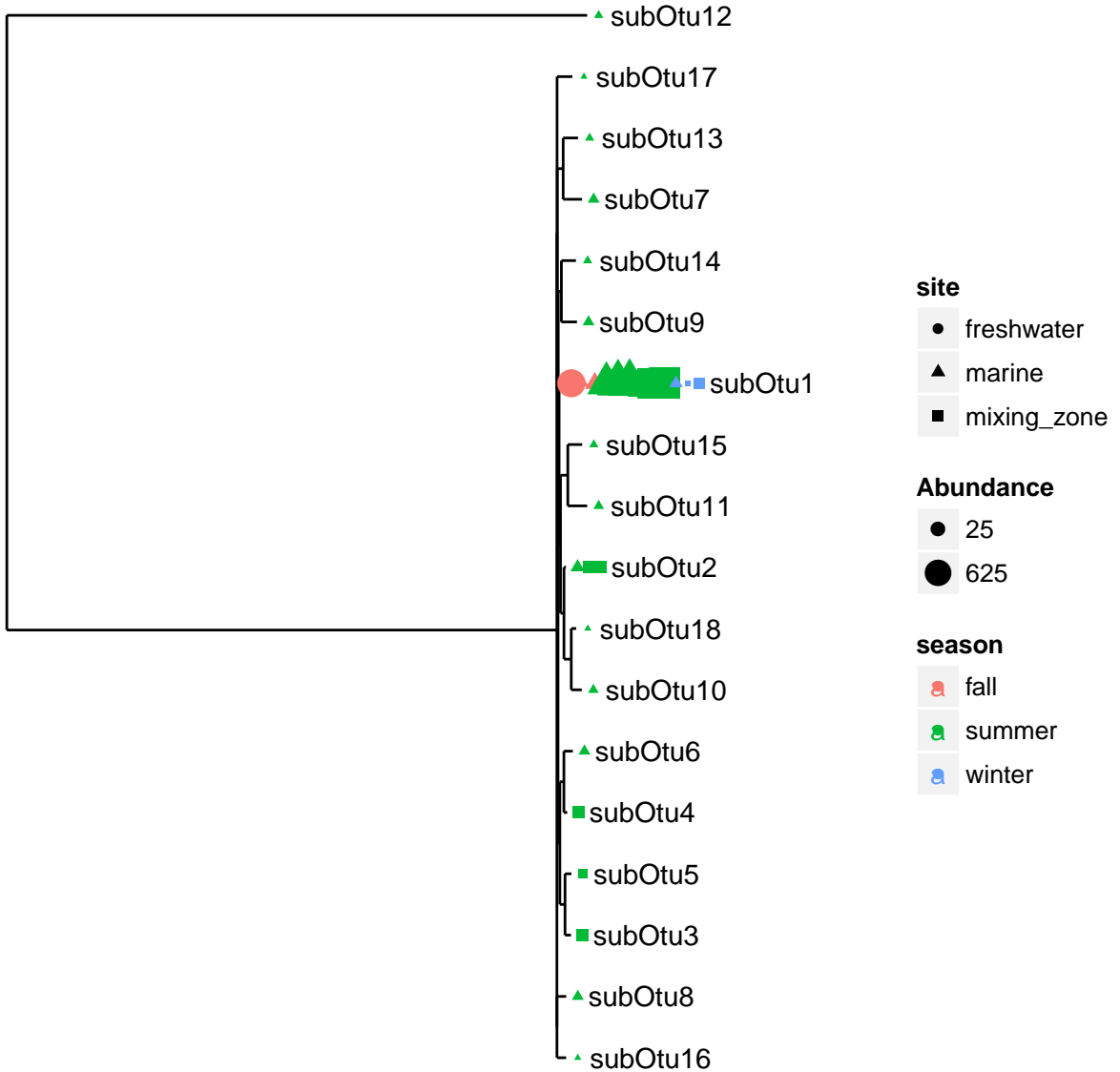
**season**

- fall
- spring
- summer
- winter

**Abundance**

- 1
- 25
- 625

otu11



otu12

subOtu2

**site**

- freshwater
- ▲ marine
- mixing\_zone

**season**

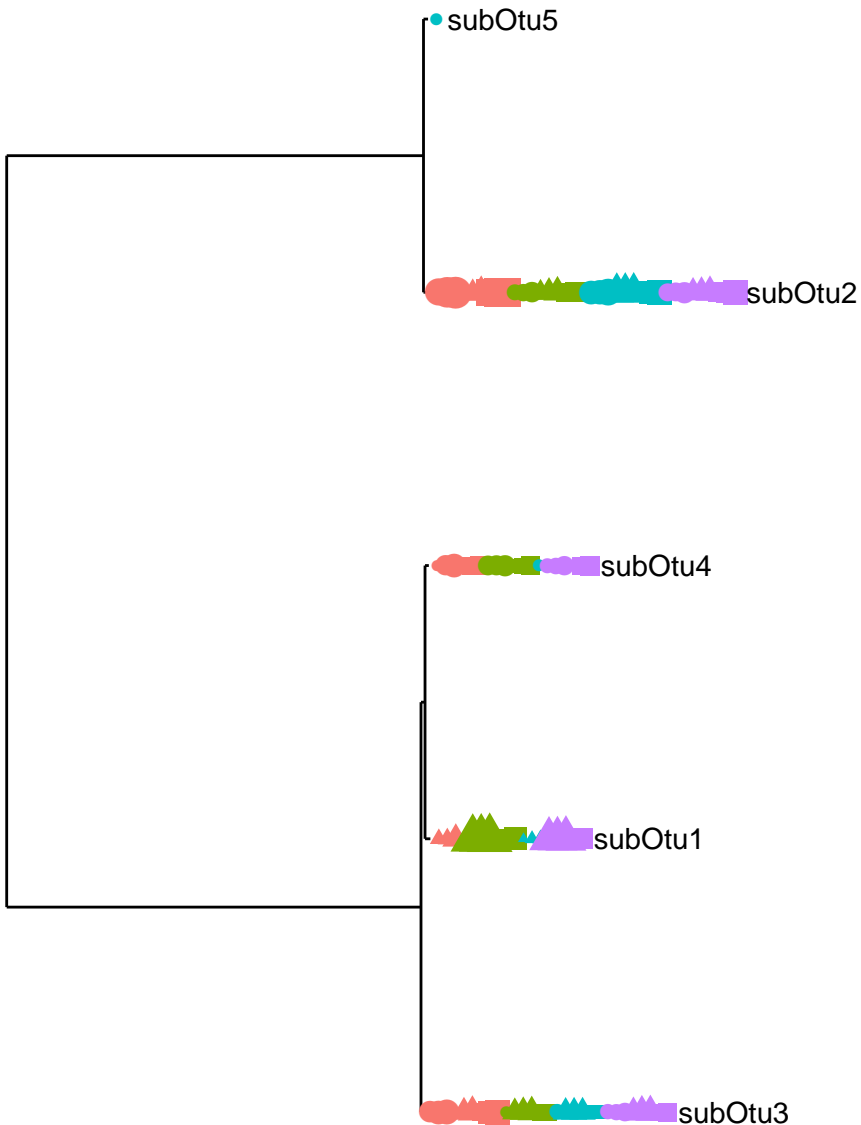
- fall
- spring
- summer
- winter

**Abundance**

- 25
- 625

subOtu1

otu13



**site**

- freshwater
- ▲ marine
- mixing\_zone

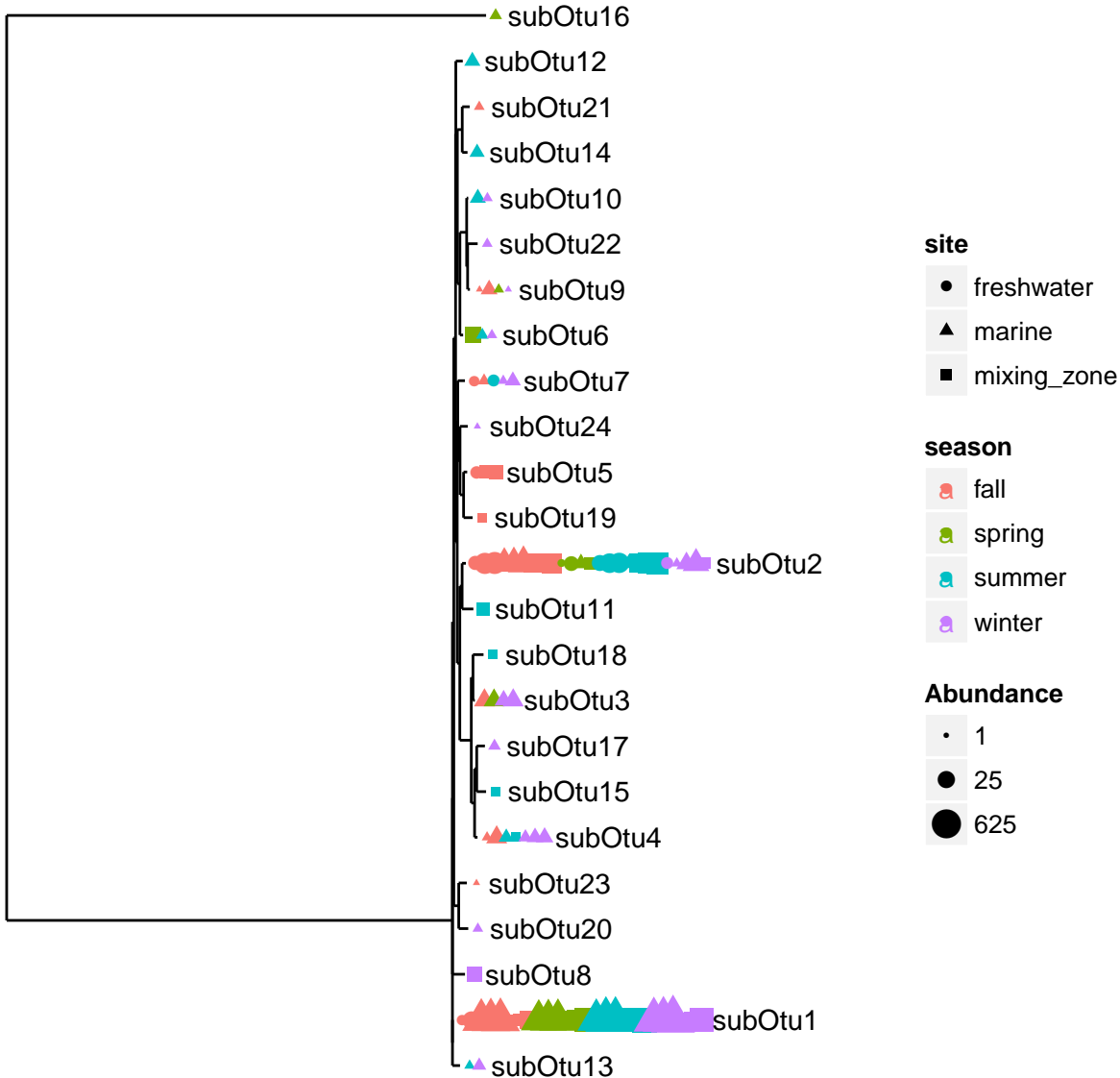
**season**

- fall
- spring
- summer
- winter

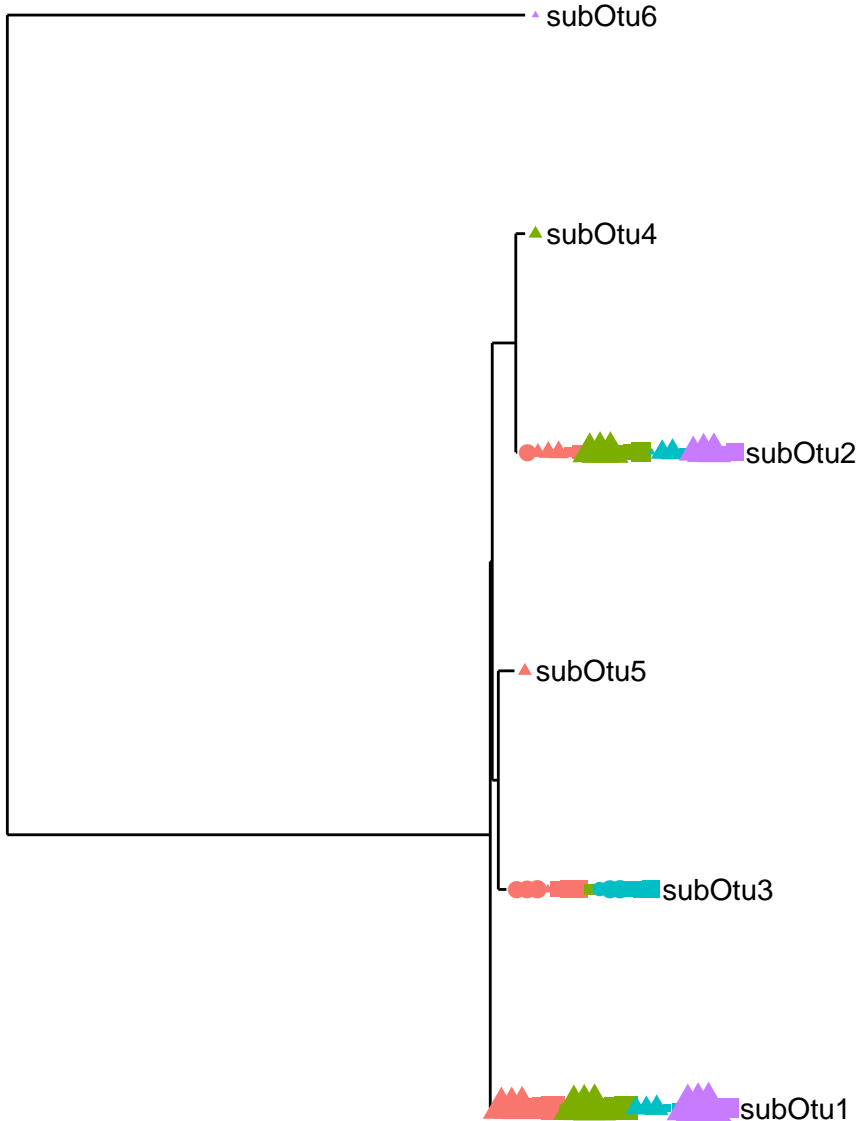
**Abundance**

- 1
- 25
- 625

otu14



otu15



**site**

- freshwater
- ▲ marine
- mixing\_zone

**season**

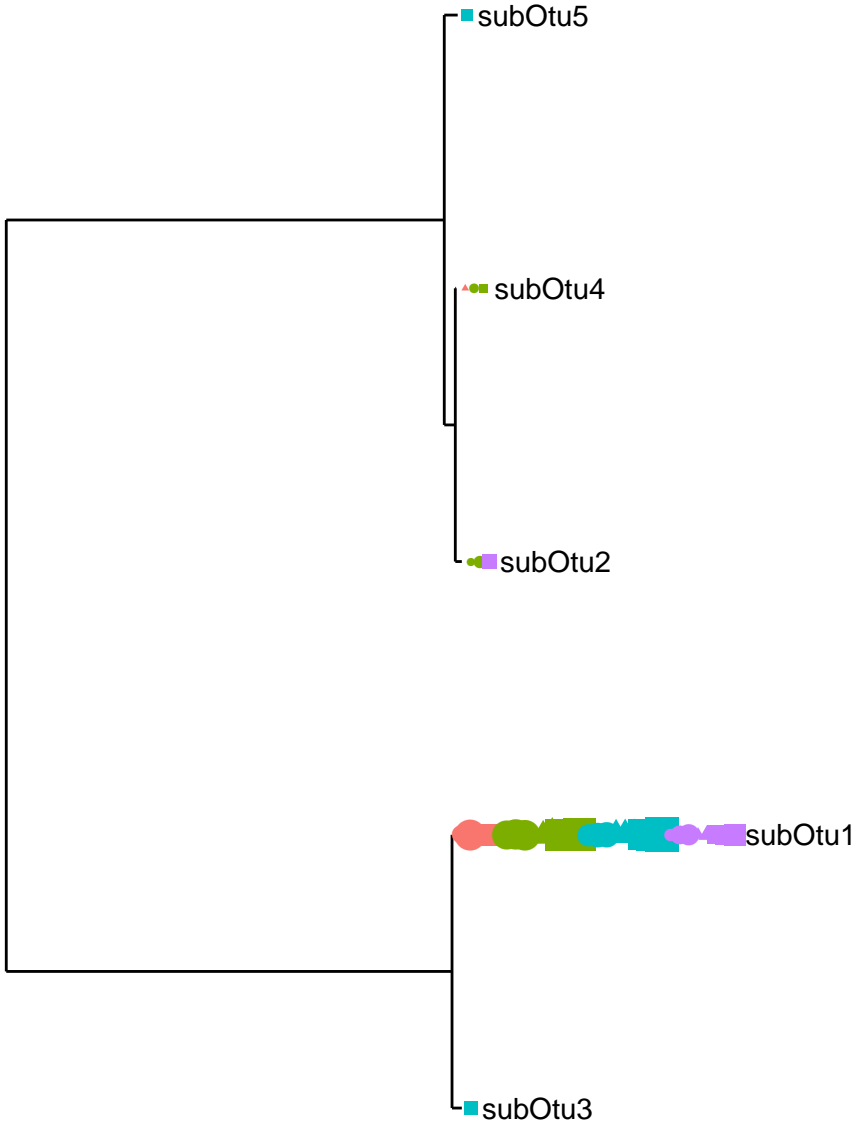
- fall
- spring
- summer
- winter

**Abundance**

- 25
- 625



otu16



**site**

- freshwater
- ▲ marine
- mixing\_zone

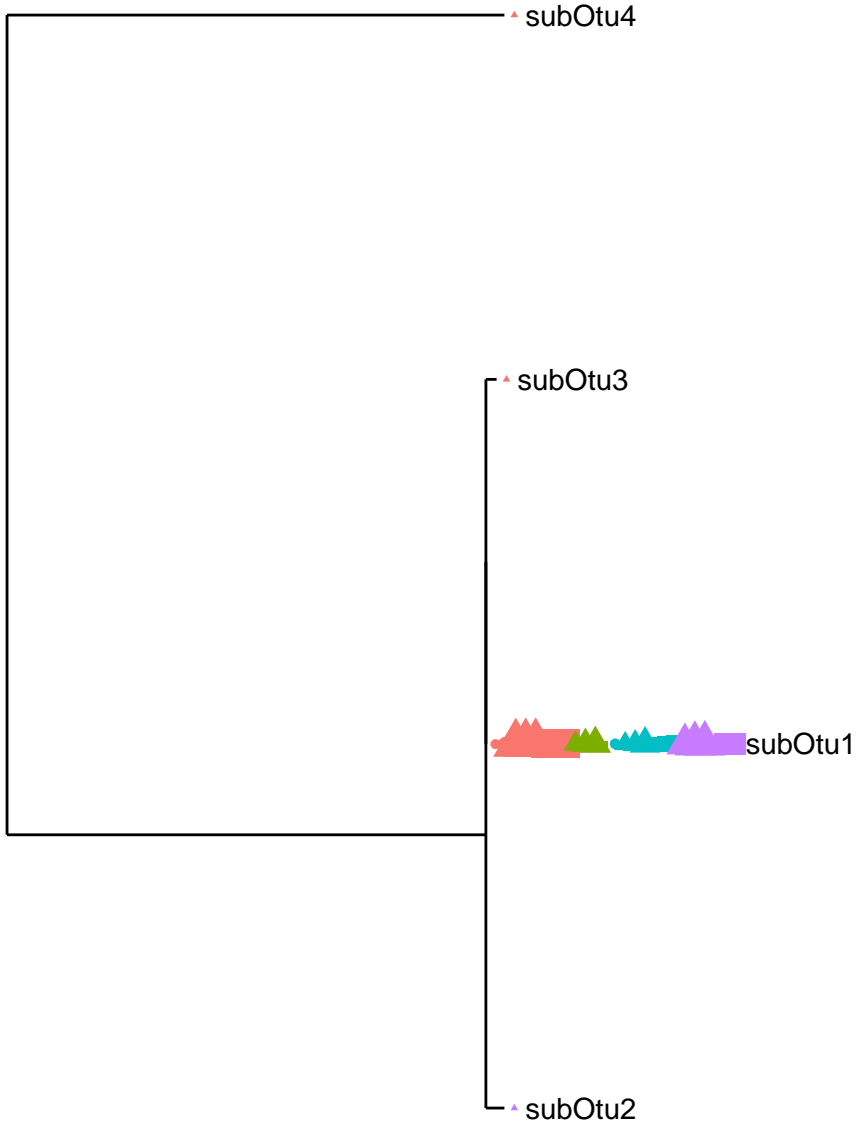
**season**

- fall
- spring
- summer
- winter

**Abundance**

- 1
- 25
- 625

otu17



**site**

- freshwater
- ▲ marine
- mixing\_zone

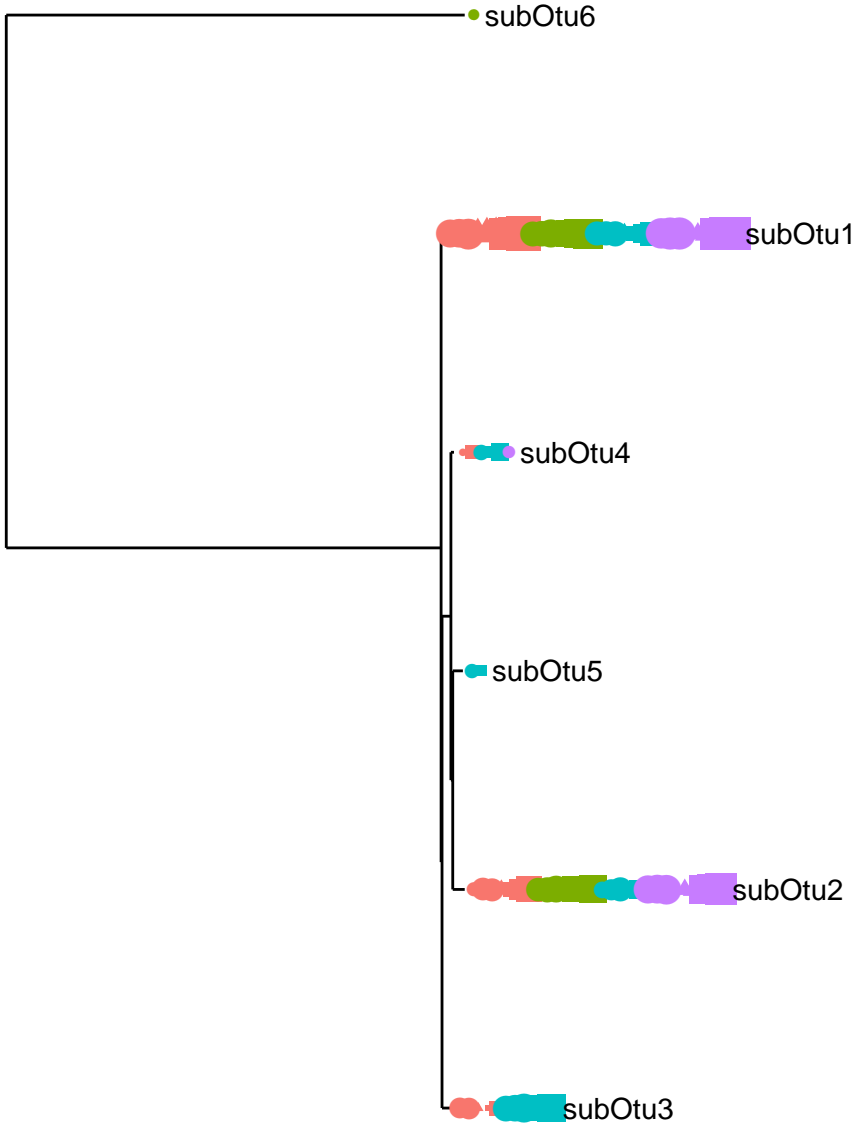
**season**

- fall
- spring
- summer
- winter

**Abundance**

- 25
- 625

otu18



**site**

- freshwater
- ▲ marine
- mixing\_zone

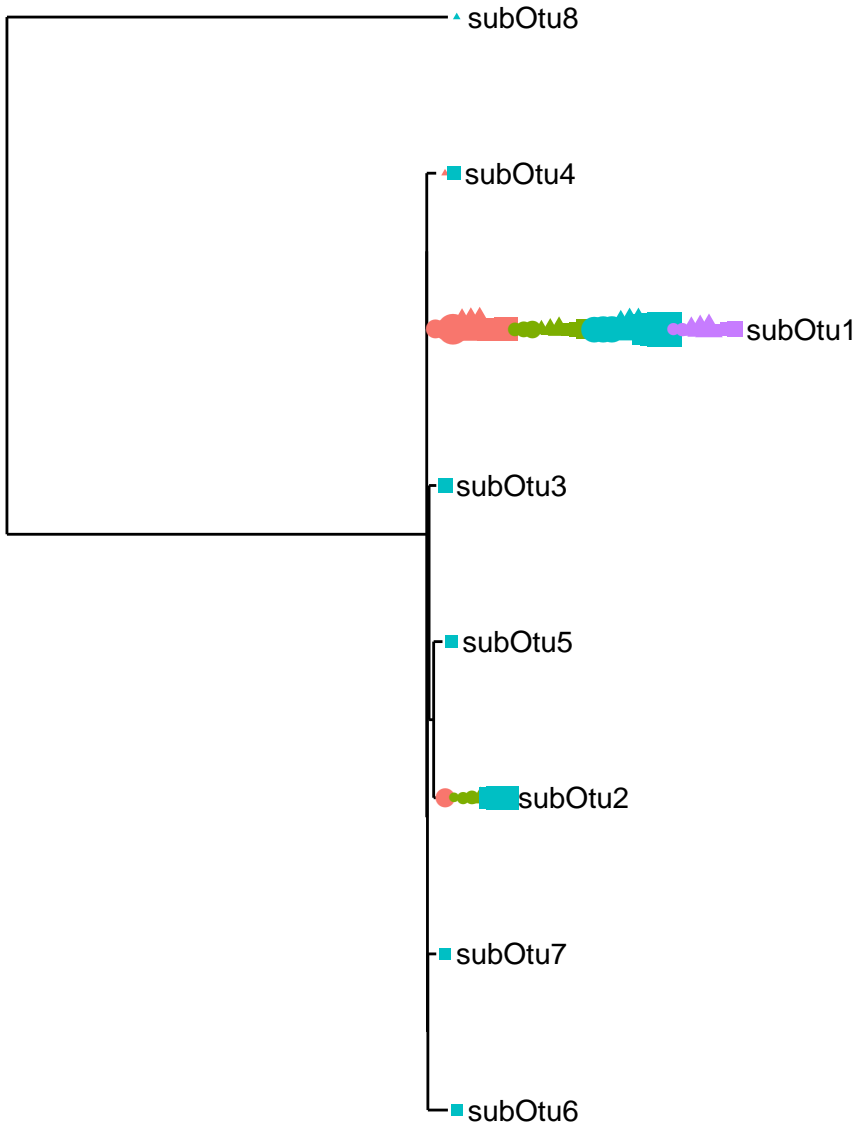
**season**

- fall
- spring
- summer
- winter

**Abundance**

- 5
- 25
- 125

otu19



**site**

- freshwater
- ▲ marine
- mixing\_zone

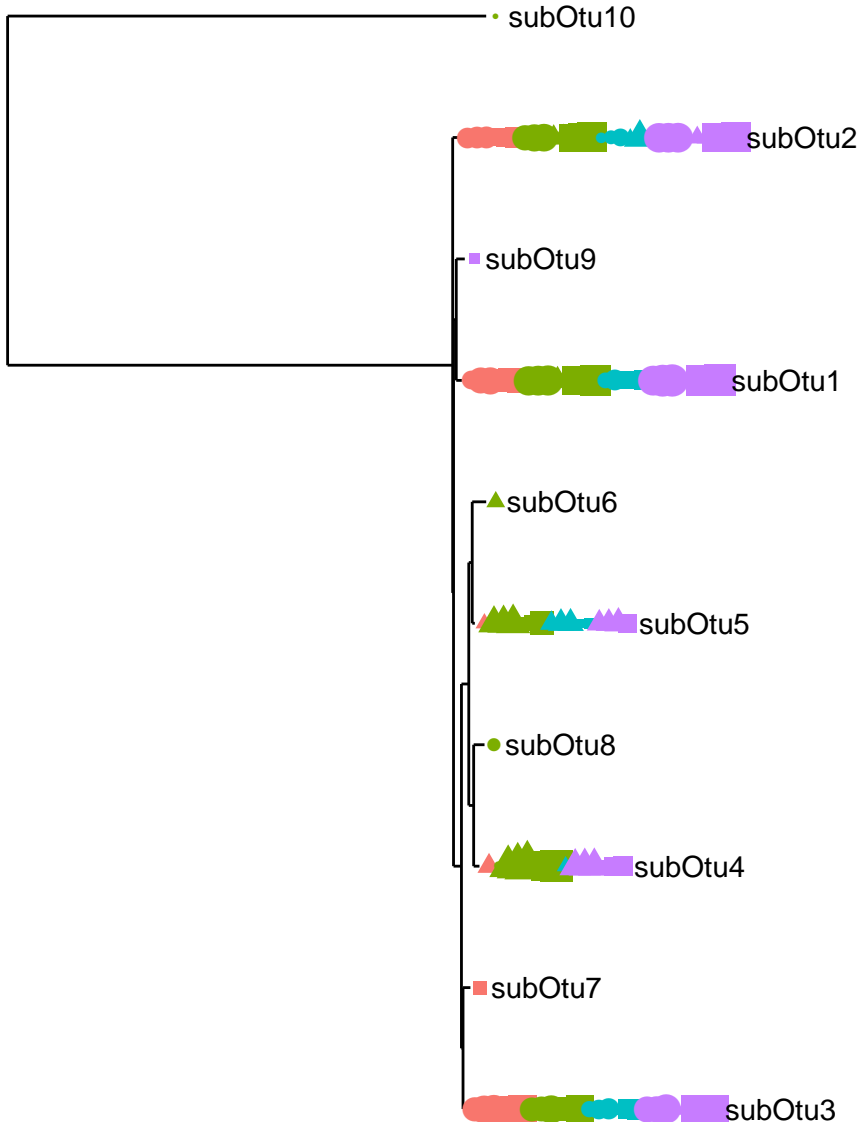
**season**

- fall
- spring
- summer
- winter

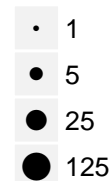
**Abundance**

- 1
- 25
- 625

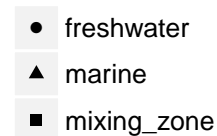
otu20



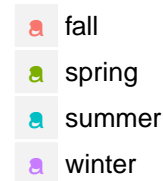
**Abundance**



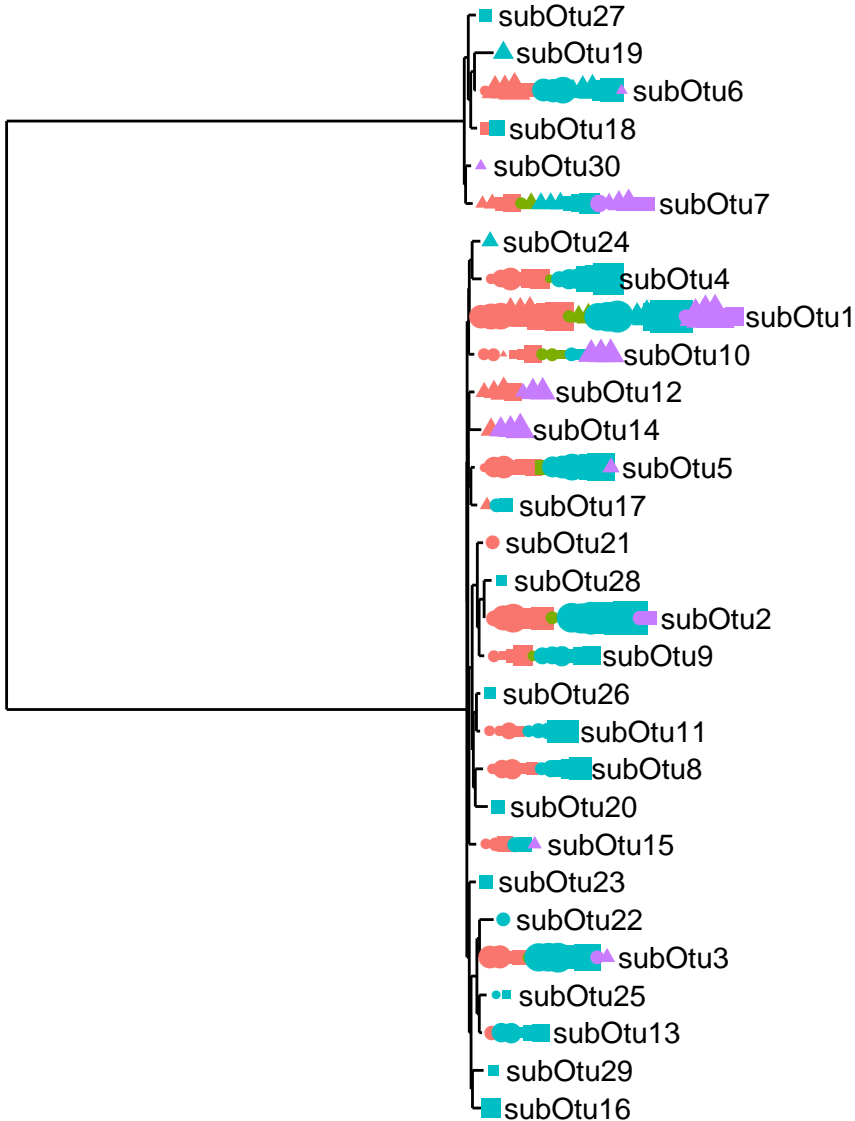
**site**



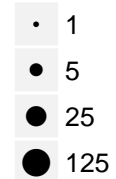
**season**



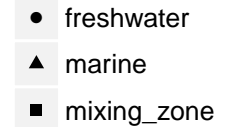
otu21



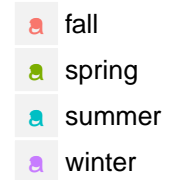
**Abundance**



**site**



**season**



otu22

• subOtu3

■ subOtu2

● subOtu1

**site**

- freshwater
- ▲ marine
- mixing\_zone

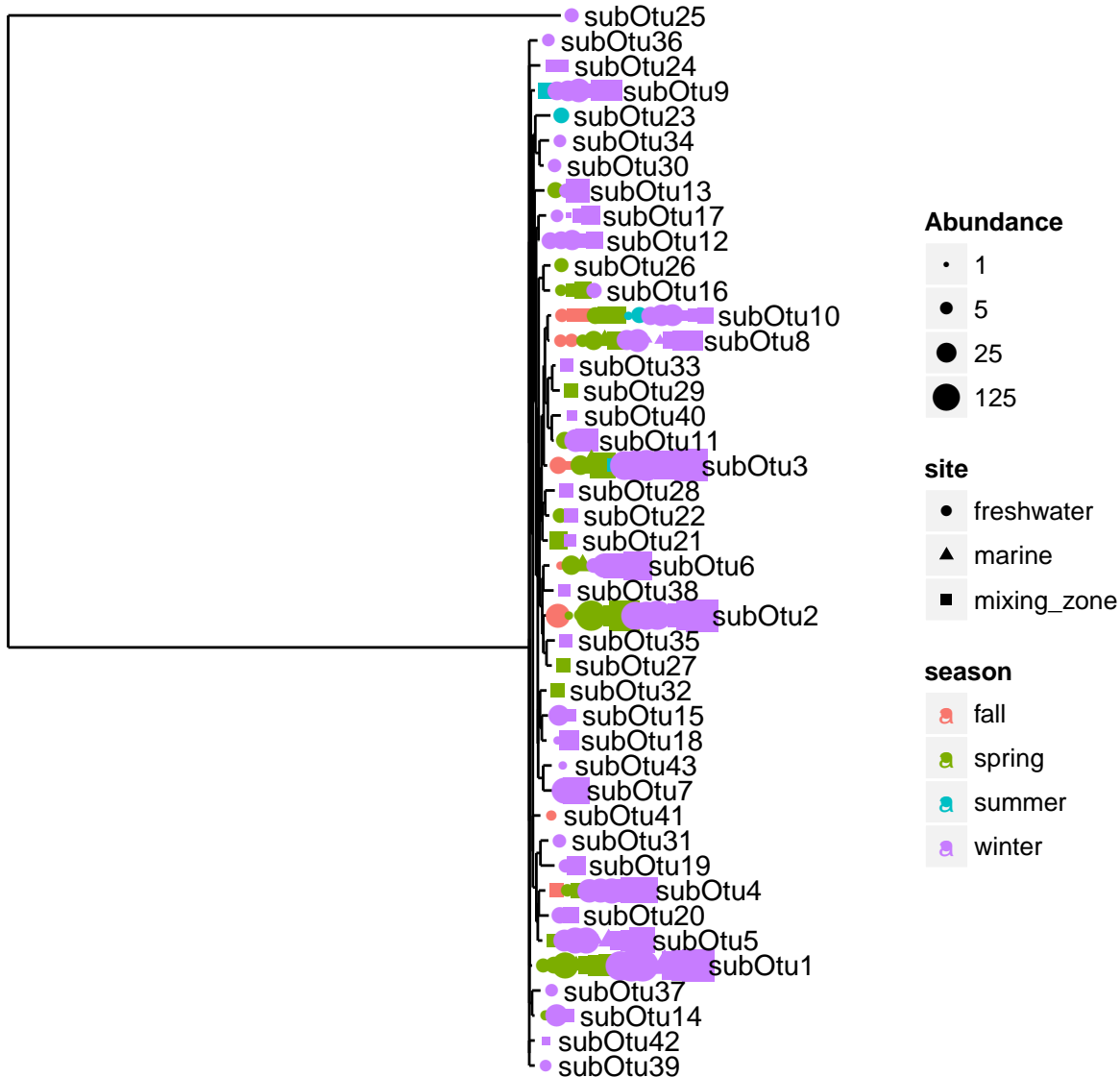
**season**

- fall
- spring
- summer
- winter

**Abundance**

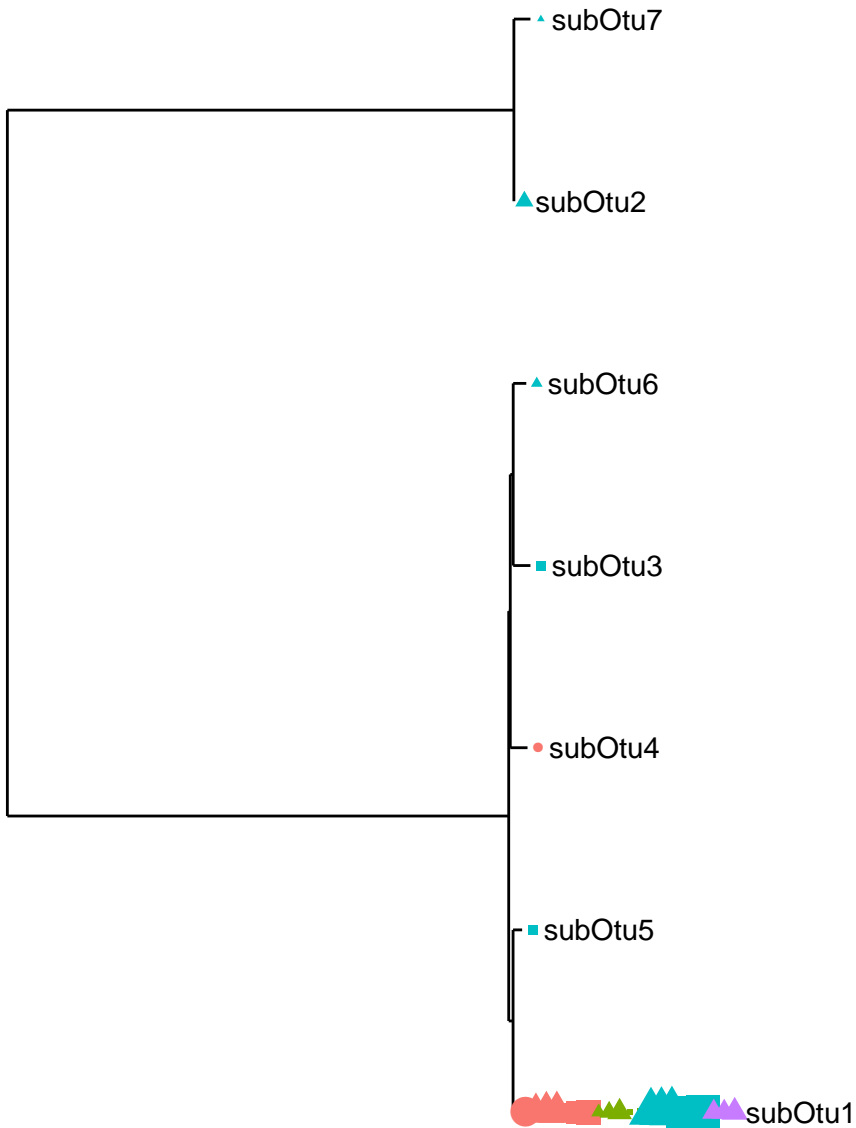
- 1
- 25
- 625

otu23





otu24



**site**

- freshwater
- ▲ marine
- mixing\_zone

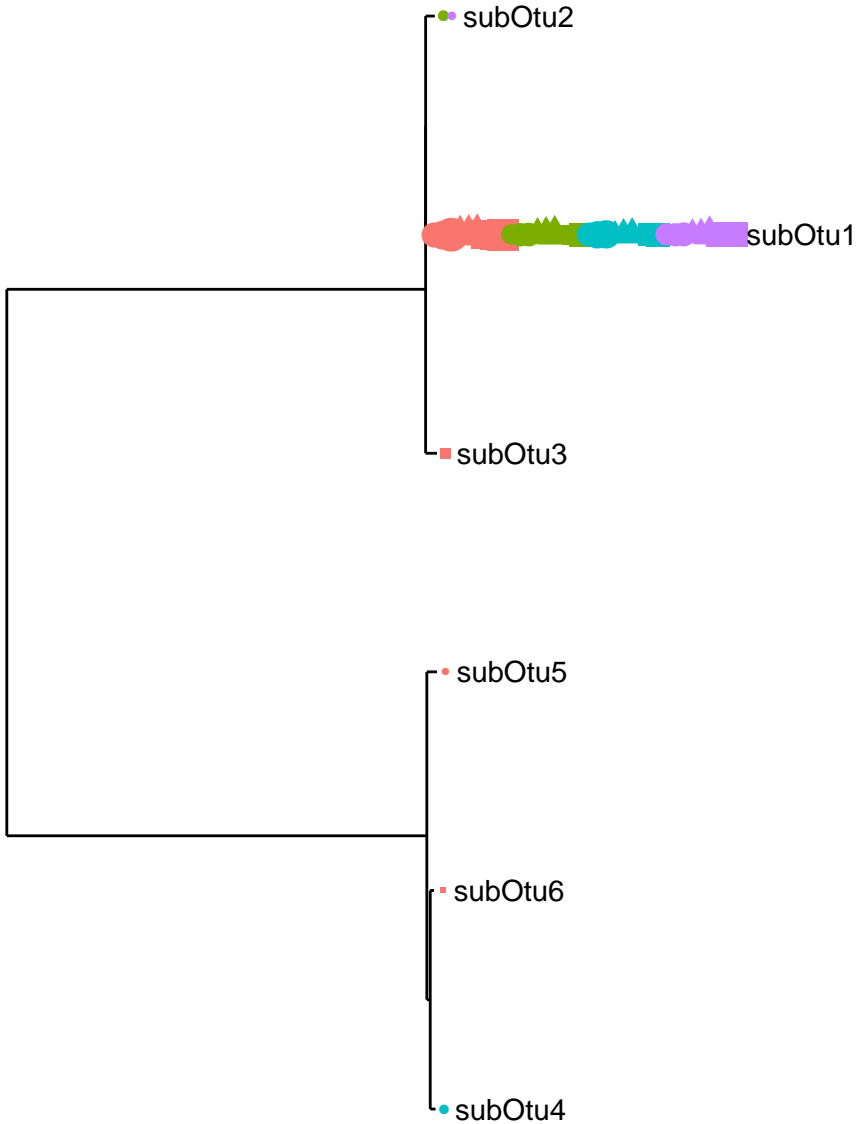
**season**

- fall
- spring
- summer
- winter

**Abundance**

- 25
- 625

otu25



**site**

- freshwater
- ▲ marine
- mixing\_zone

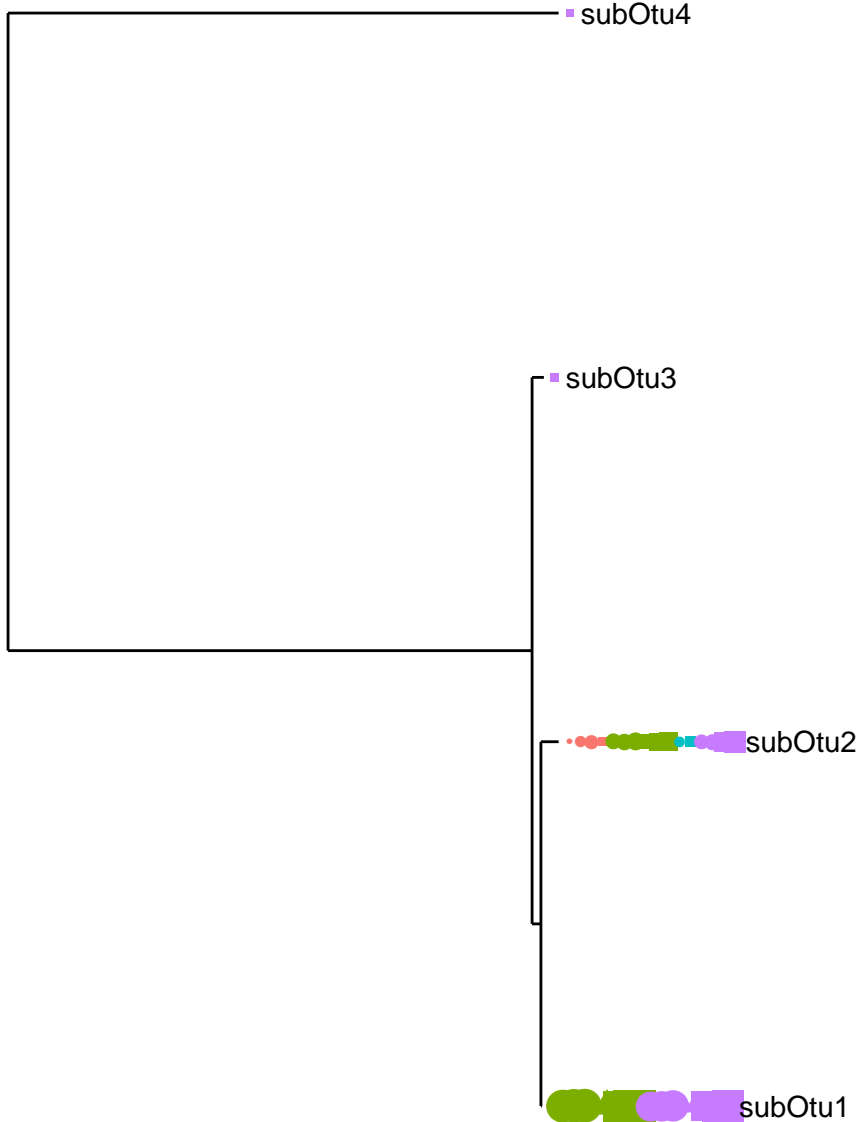
**season**

- fall
- spring
- summer
- winter

**Abundance**

- 25
- 625

otu26



**site**

- freshwater
- ▲ marine
- mixing\_zone

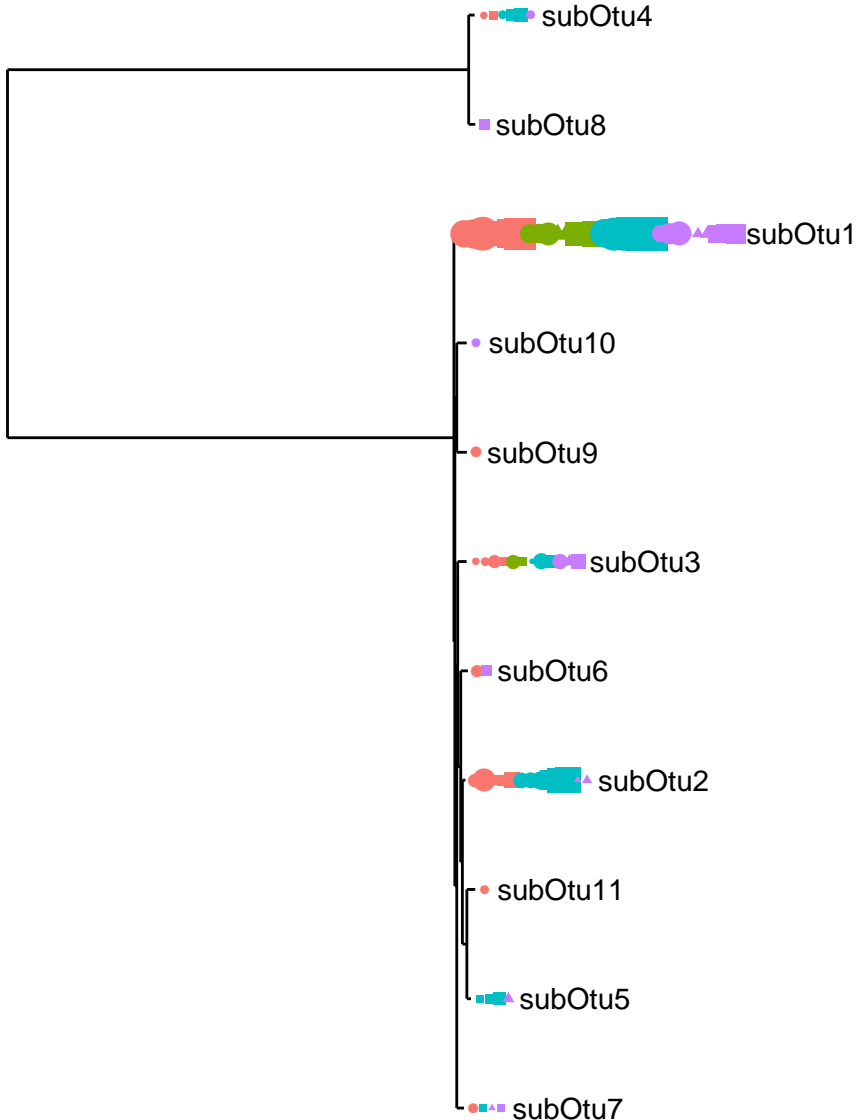
**season**

- fall
- spring
- summer
- winter

**Abundance**

- 25
- 625

otu27



**site**

- freshwater
- ▲ marine
- mixing\_zone

**season**

- fall
- spring
- summer
- winter

**Abundance**

- 25
- 625

otu28

▲ subOtu3

● subOtu2

subOtu1

**site**

- freshwater
- ▲ marine
- mixing\_zone

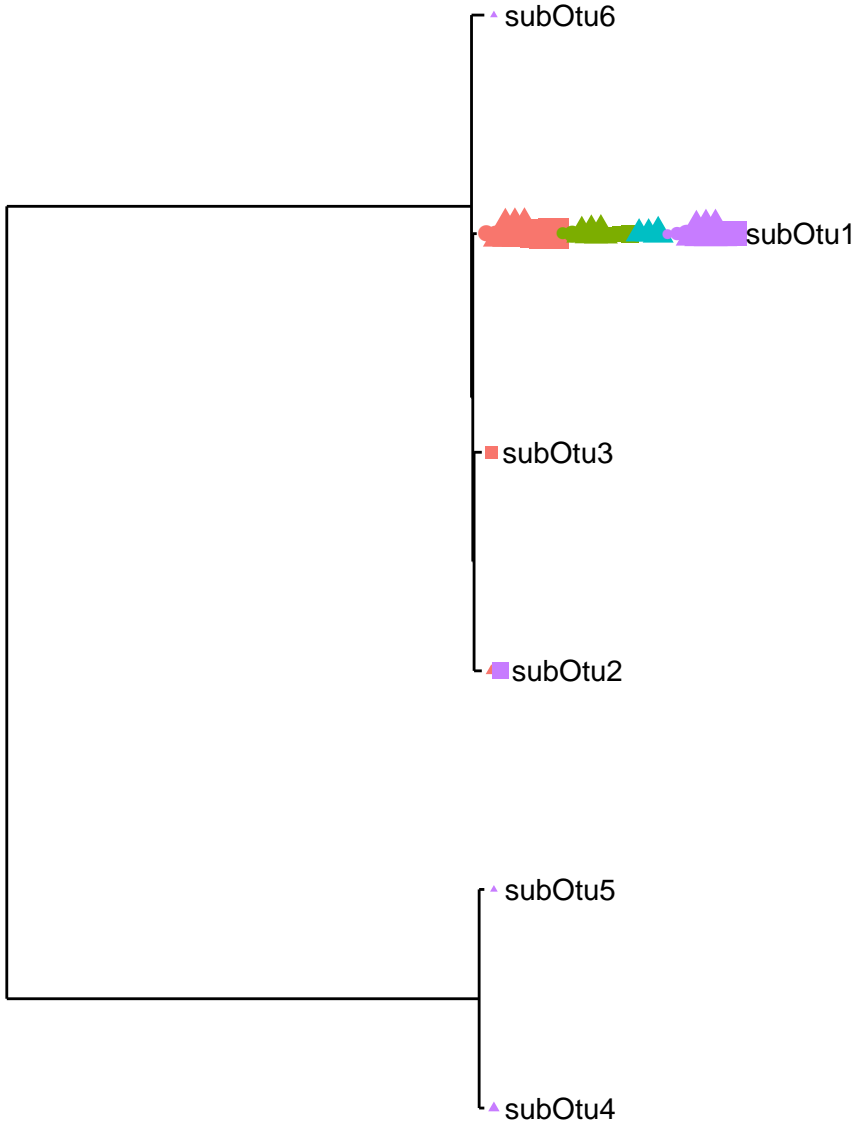
**season**

- fall
- spring
- summer
- winter

**Abundance**

- 25
- 625

otu29



**site**

- freshwater
- ▲ marine
- mixing\_zone

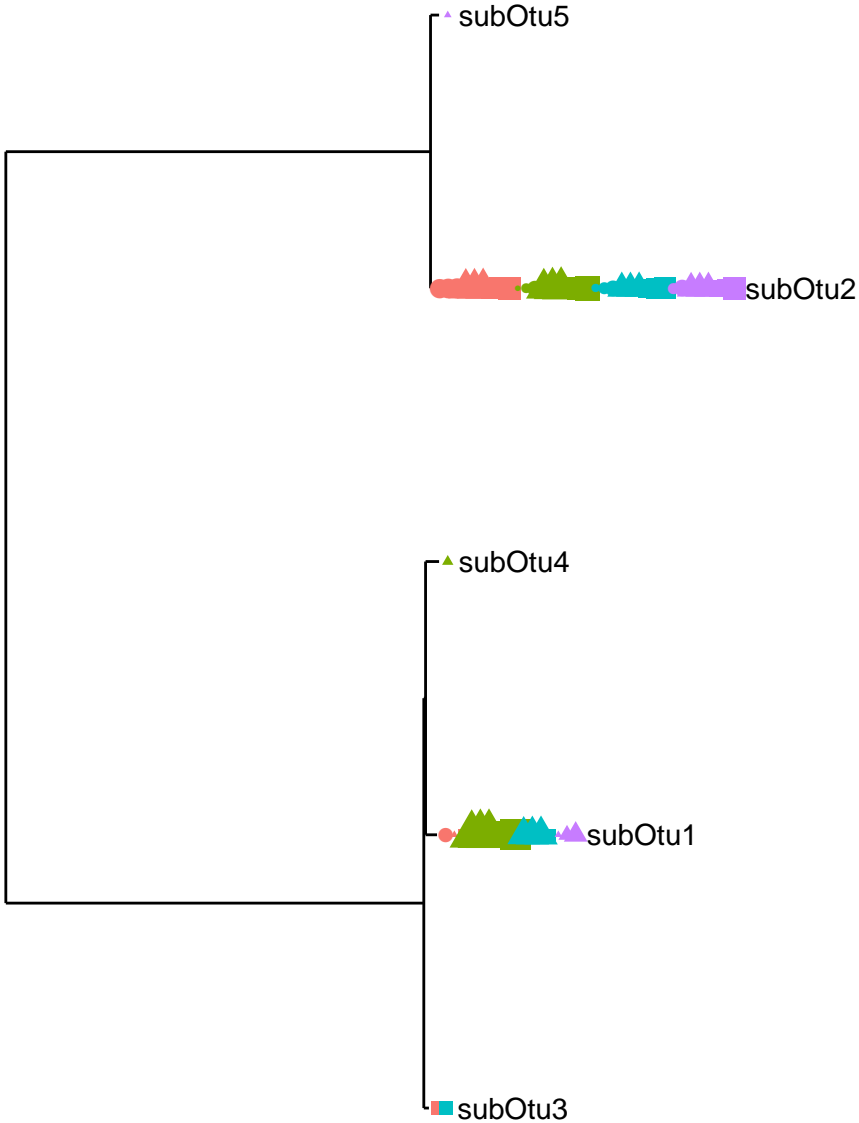
**season**

- fall
- spring
- summer
- winter

**Abundance**

- 1
- 25
- 625

otu30



**site**

- freshwater
- ▲ marine
- mixing\_zone

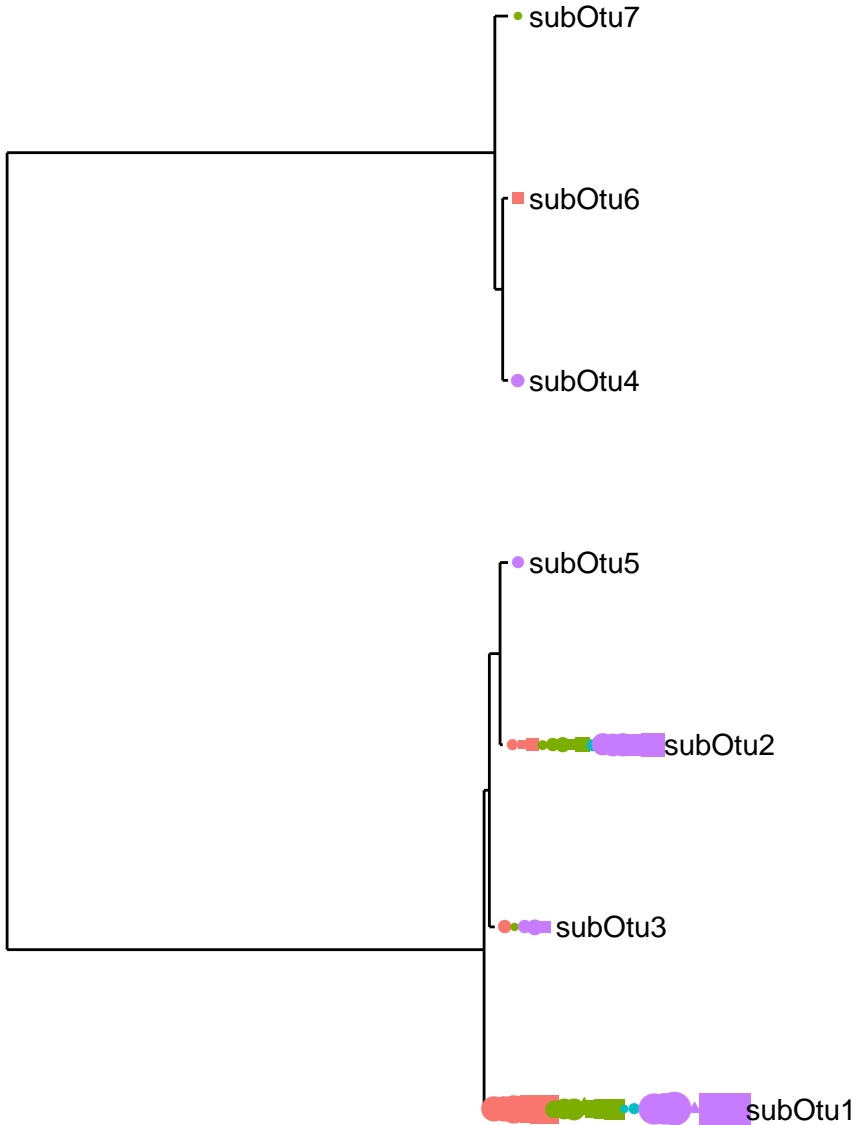
**season**

- fall
- spring
- summer
- winter

**Abundance**

- 1
- 25
- 625

otu32



**site**

- freshwater
- ▲ marine
- mixing\_zone

**season**

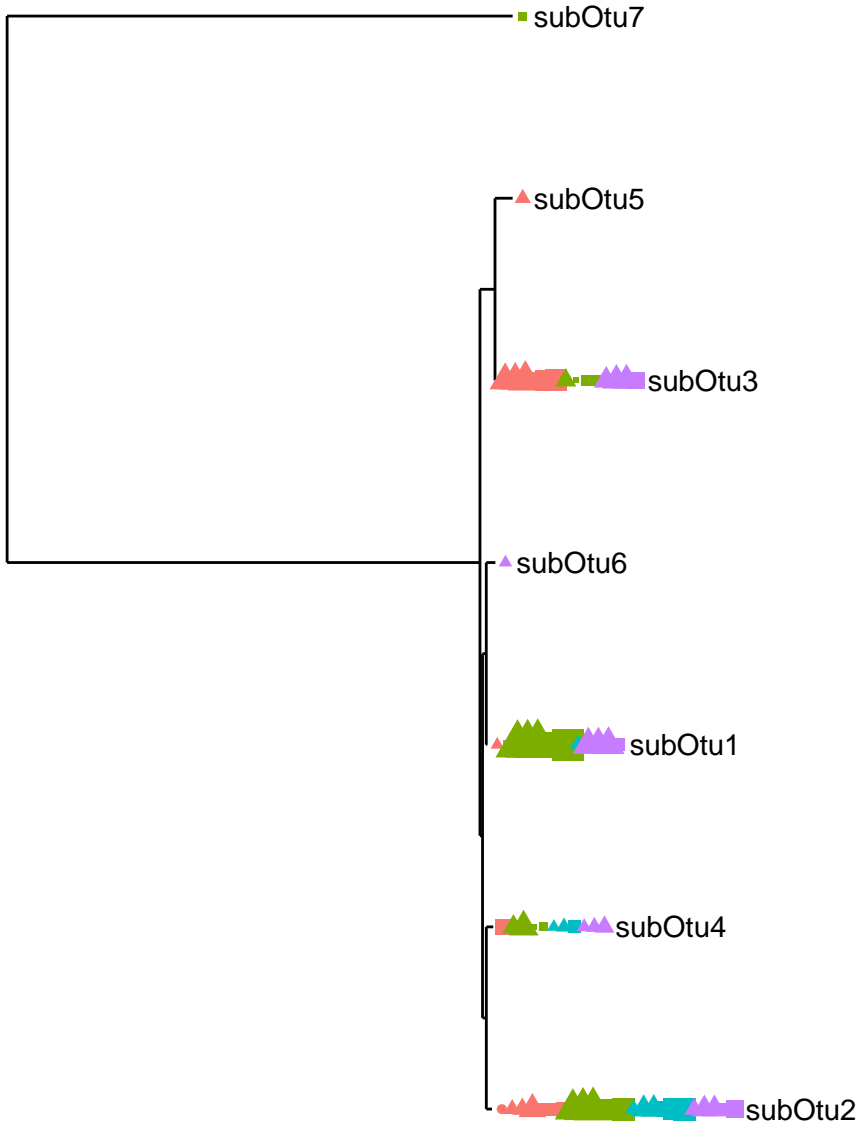
- fall
- spring
- summer
- winter

**Abundance**

- 1
- 25
- 625



otu33



**site**

- freshwater
- ▲ marine
- mixing\_zone

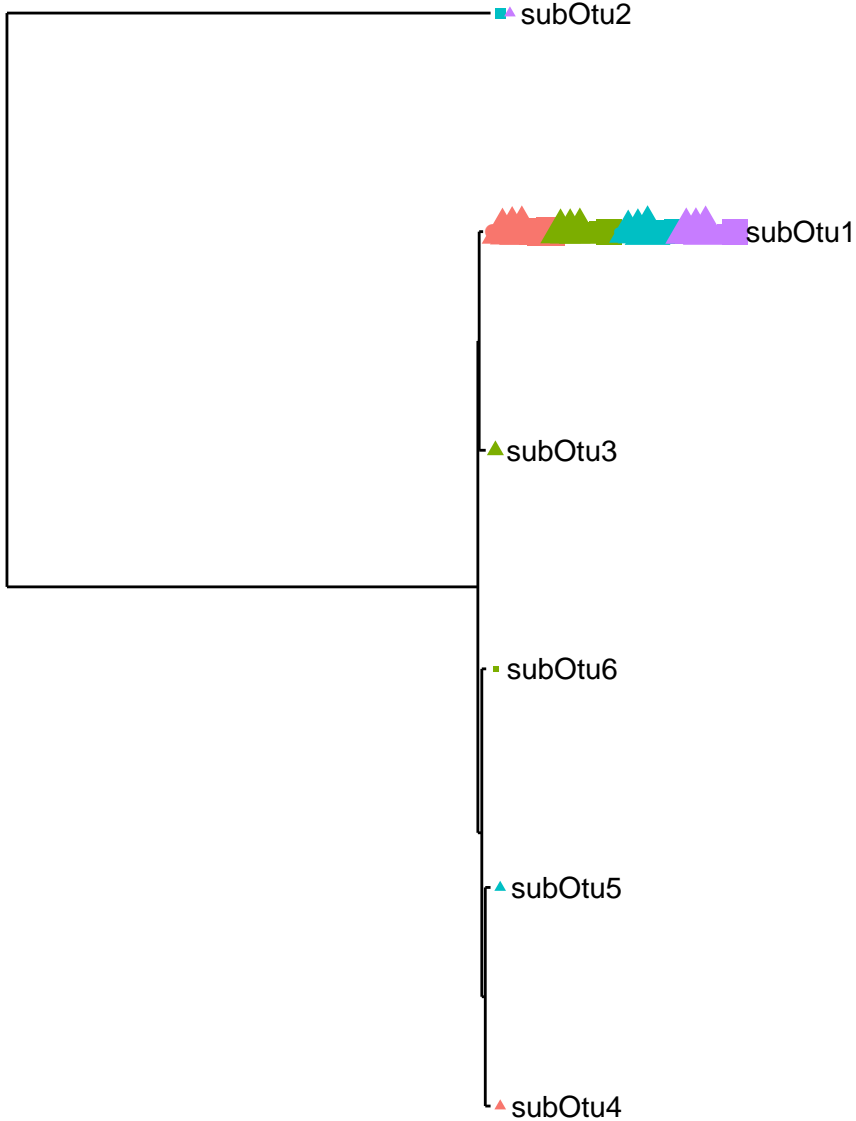
**season**

- fall
- spring
- summer
- winter

**Abundance**

- 1
- 25
- 625

otu34



**site**

- freshwater
- ▲ marine
- mixing\_zone

**season**

- fall
- spring
- summer
- winter

**Abundance**

- 1
- 25
- 625

otu36



**site**

- freshwater
- ▲ marine
- mixing\_zone

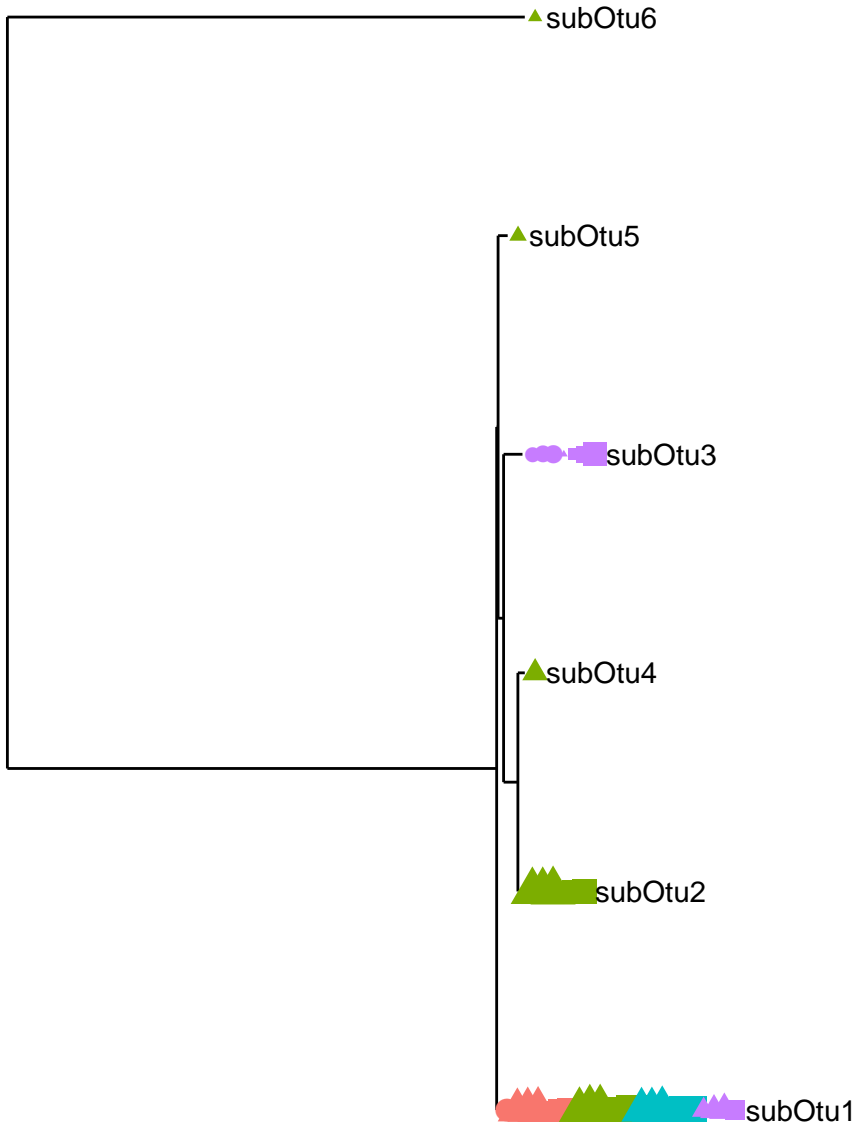
**season**

- fall
- spring
- summer
- winter

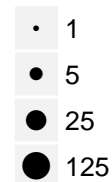
**Abundance**

- 1
- 25
- 625

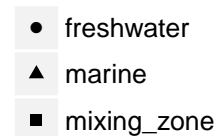
otu37



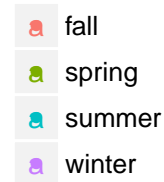
**Abundance**



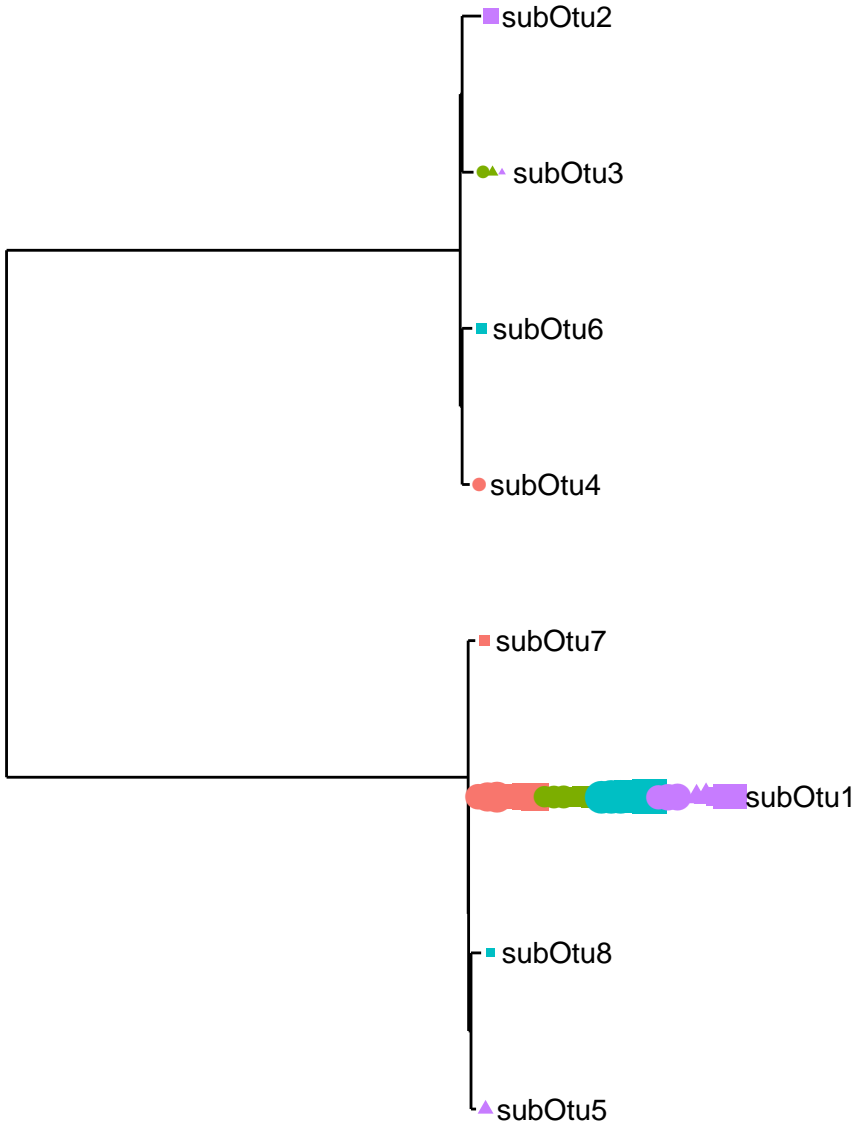
**site**



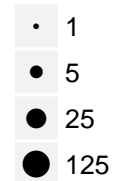
**season**



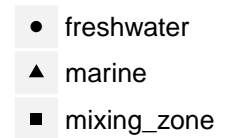
otu38



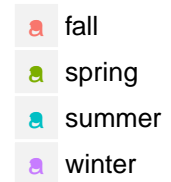
**Abundance**



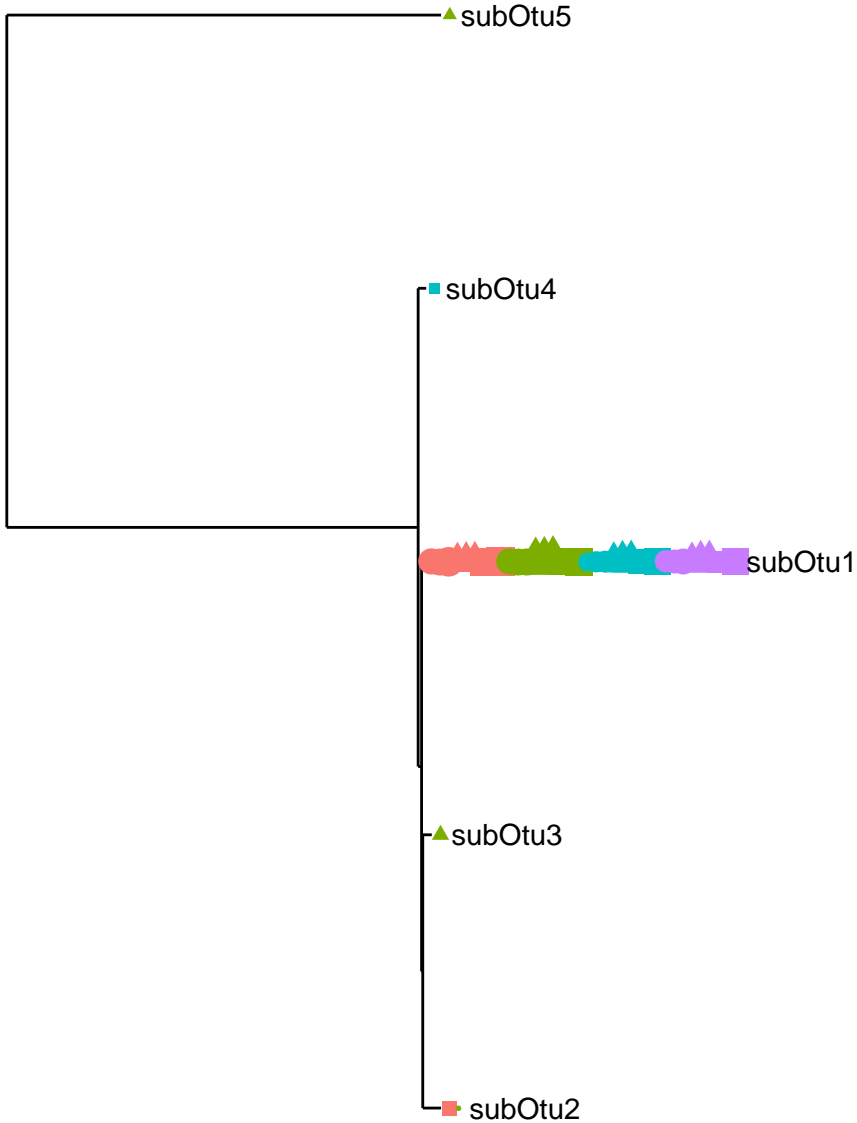
**site**



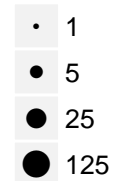
**season**



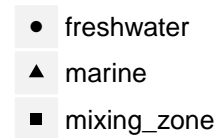
otu39



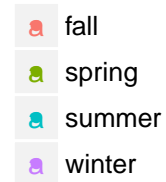
**Abundance**



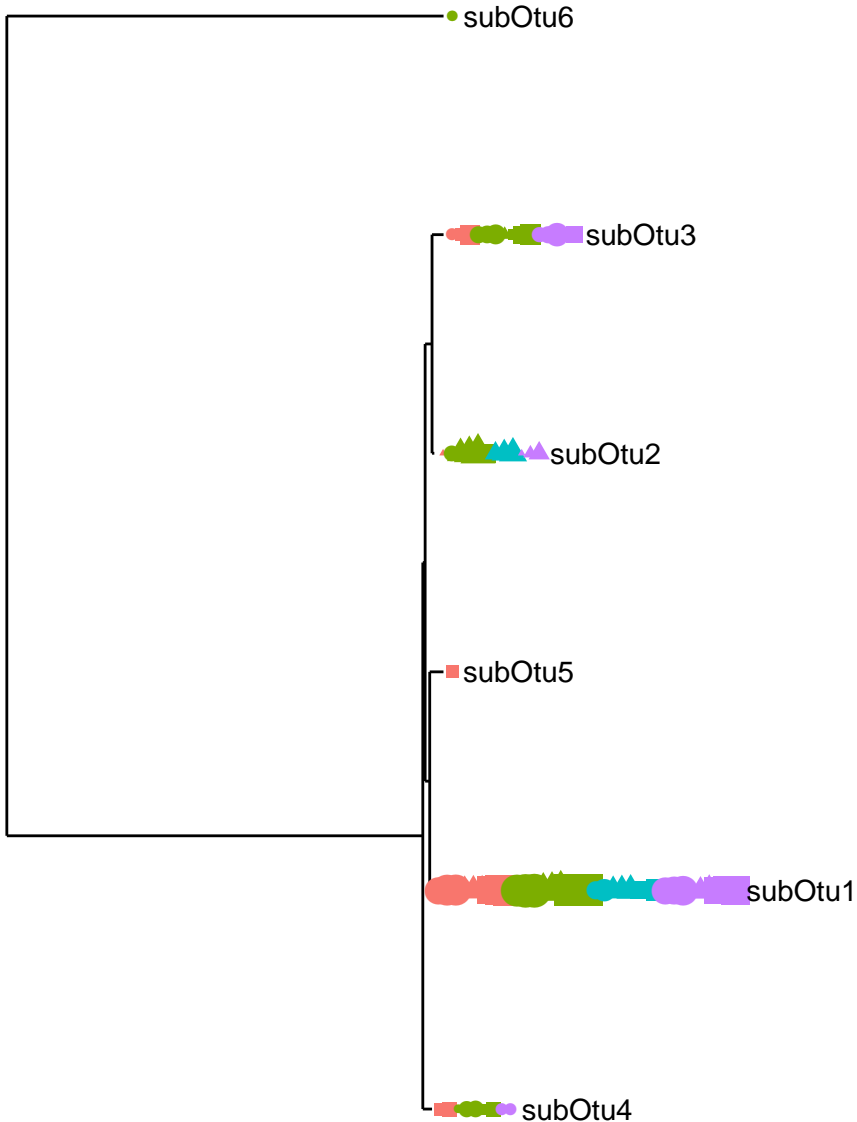
**site**



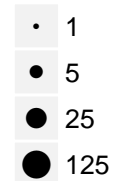
**season**



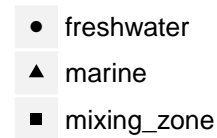
otu40



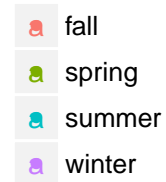
**Abundance**



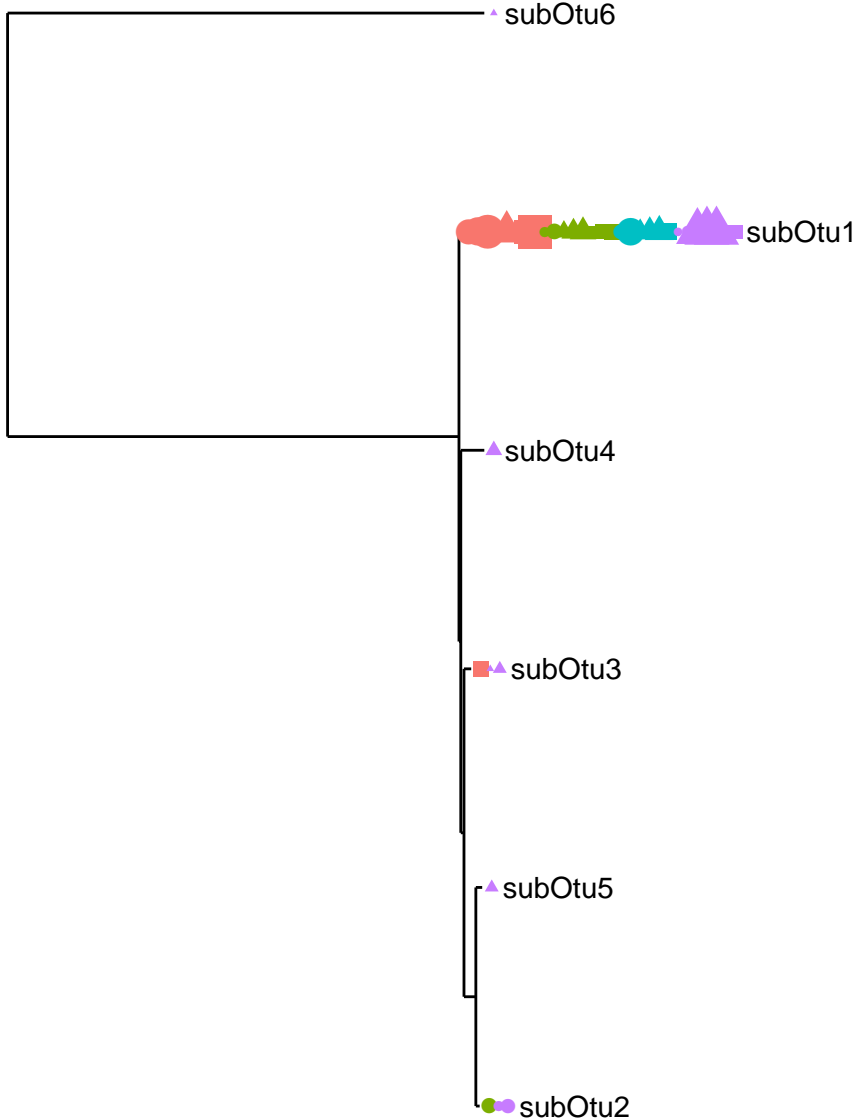
**site**



**season**



otu41



**site**

- freshwater
- ▲ marine
- mixing\_zone

**season**

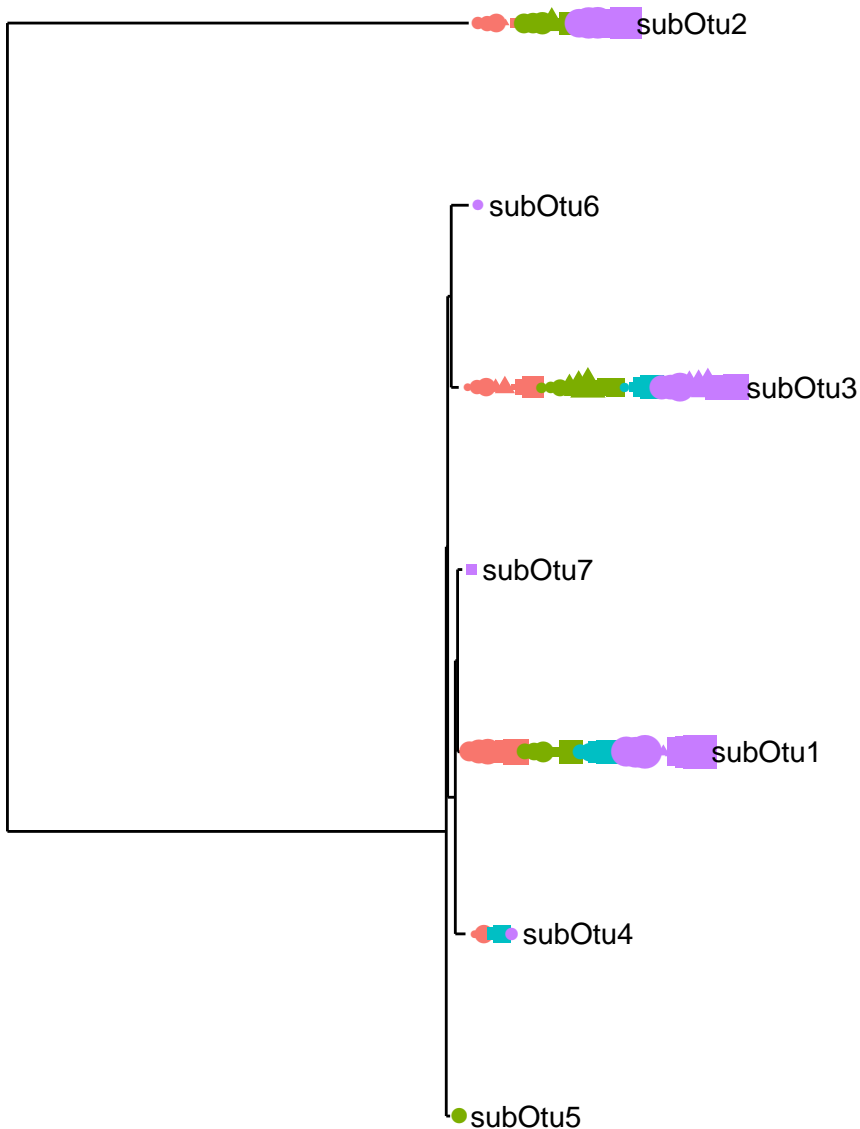
- fall
- spring
- summer
- winter

**Abundance**

- 1
- 25
- 625



otu42



**site**

- freshwater
- ▲ marine
- mixing\_zone

**season**

- fall
- spring
- summer
- winter

**Abundance**

- 5
- 25
- 125

otu43

subOtu2

**site**

- freshwater
- ▲ marine
- mixing\_zone

**season**

- fall
- spring
- summer
- winter

**Abundance**

- 25
- 625

subOtu1

otu44

▲ subOtu3

▲ subOtu2

subOtu1

**site**

- freshwater
- ▲ marine
- mixing\_zone

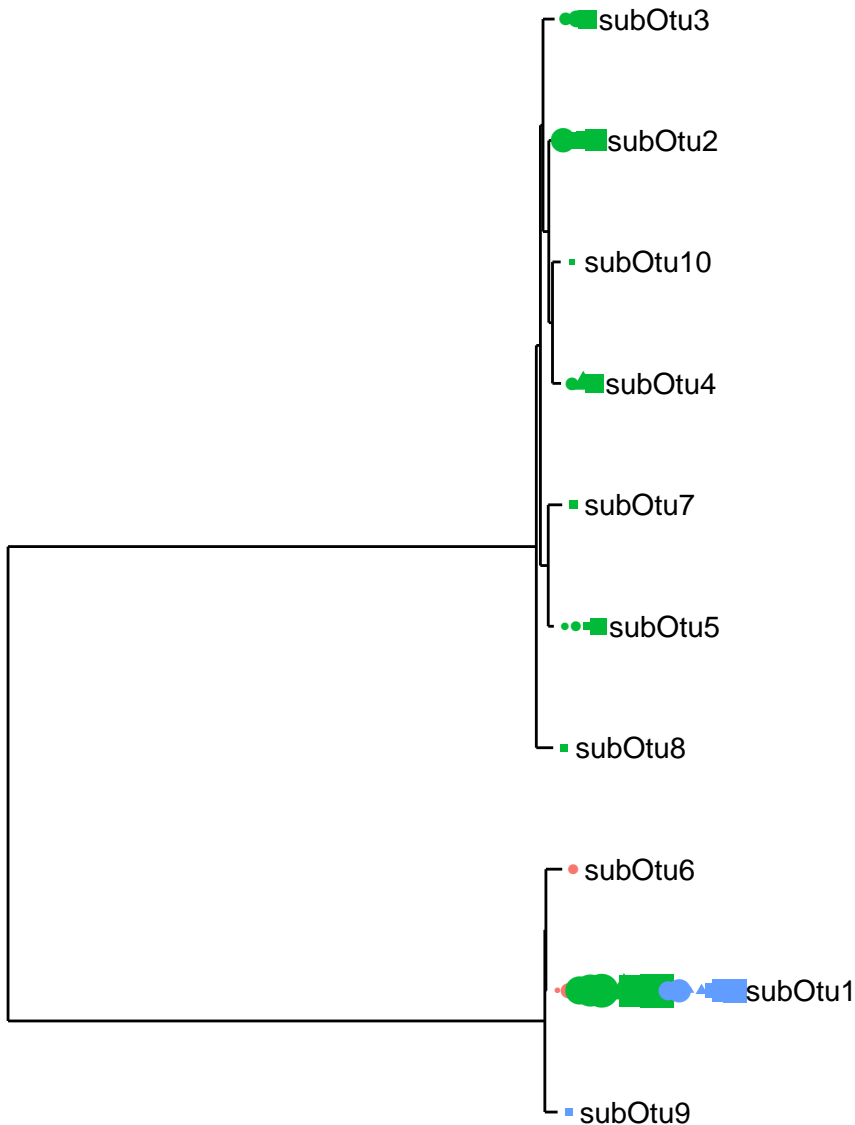
**season**

- fall
- spring
- summer
- winter

**Abundance**

- 5
- 25
- 125

otu45



otu46

• subOtu3

■ subOtu2

subOtu1

**site**

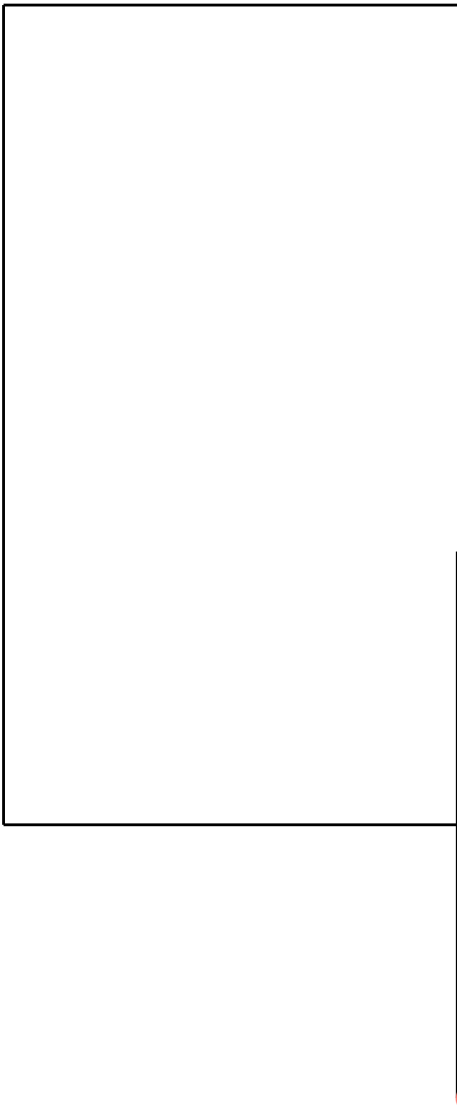
- freshwater
- ▲ marine
- mixing\_zone

**season**

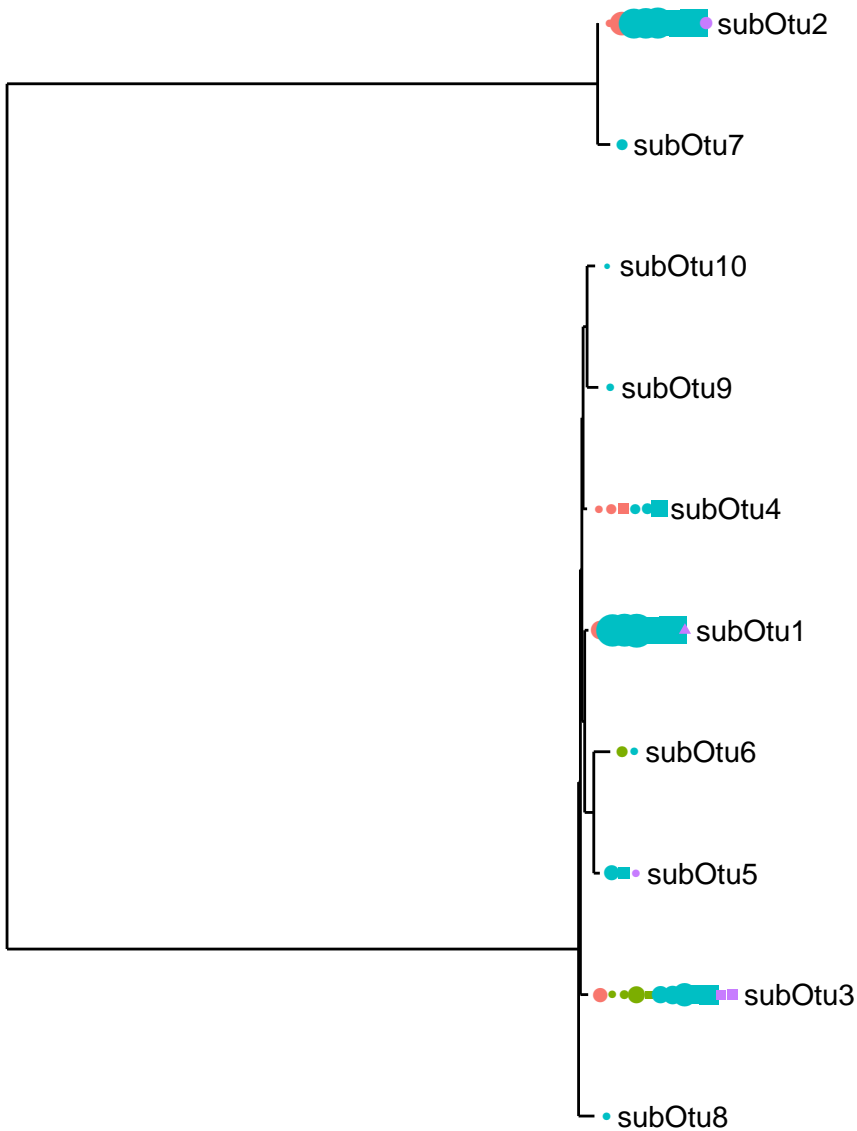
- fall
- spring
- summer
- winter

**Abundance**

- 25
- 625



otu47



**site**

- freshwater
- ▲ marine
- mixing\_zone

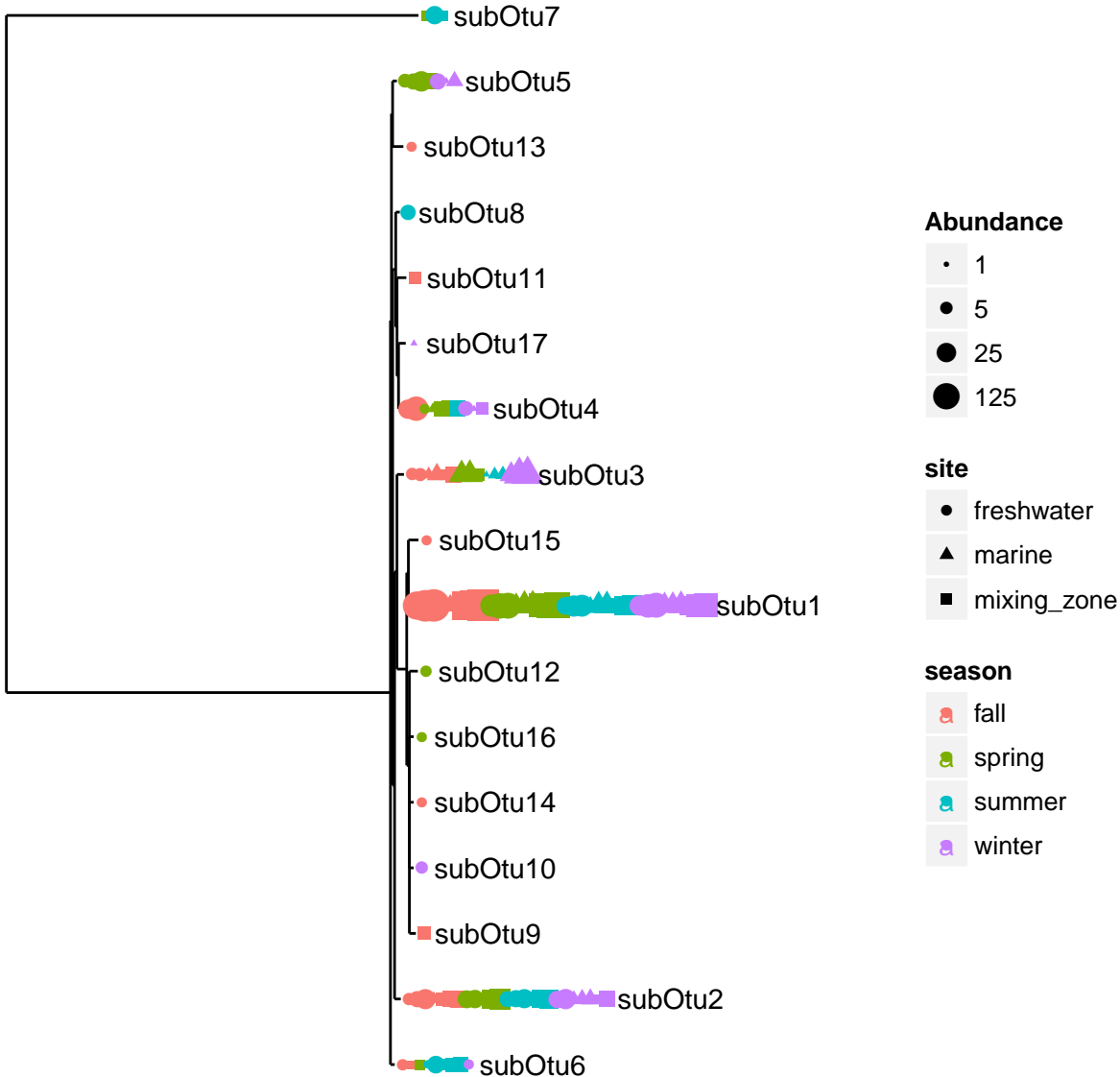
**season**

- fall
- spring
- summer
- winter

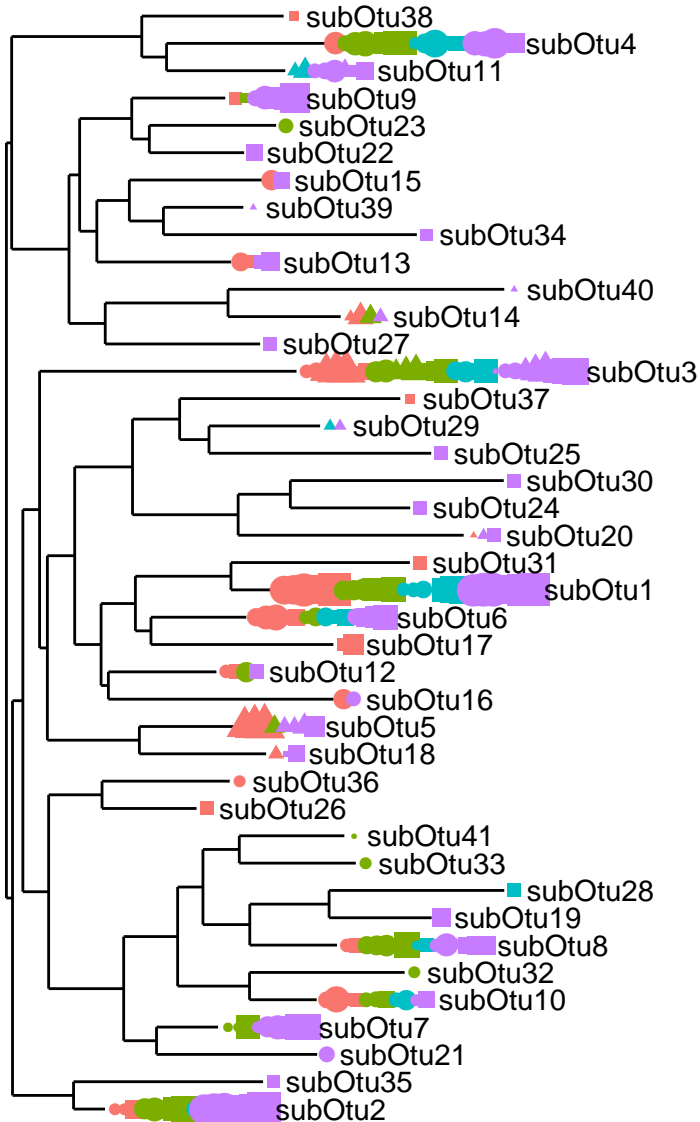
**Abundance**

- 25
- 625

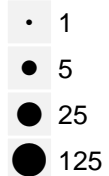
otu48



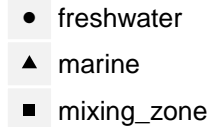
otu49



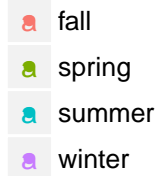
**Abundance**



**site**

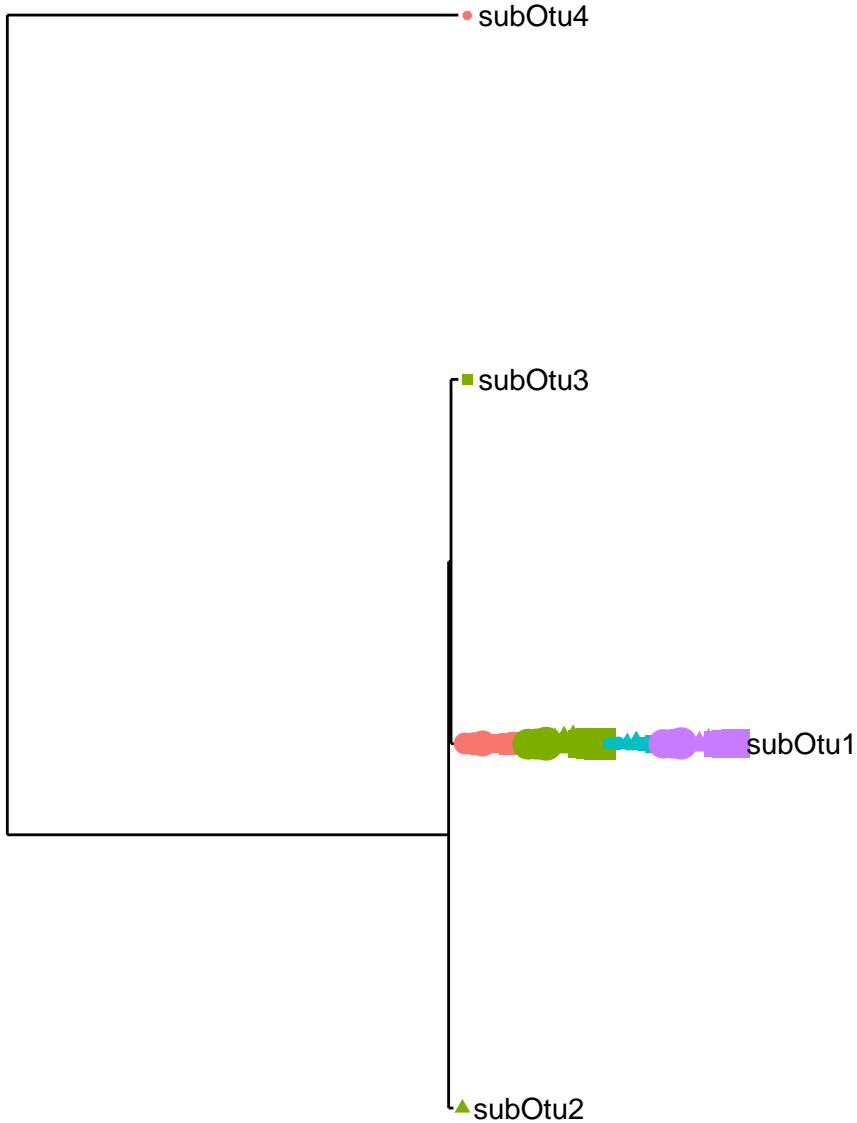


**season**





otu50



**site**

- freshwater
- ▲ marine
- mixing\_zone

**season**

- fall
- spring
- summer
- winter

**Abundance**

- 5
- 25
- 125

## 1 Methodology of 454 reads processing

2

3 In our study we applied high-throughput sequencing method (HTS) to detect phylotypes that  
4 were abundant in freshwater, and still present but rare at the brackish site, and thus to describe the  
5 microbiome of the whole estuary in more detail. We utilised 454 sequencing of the V3-V4 rRNA  
6 fragment, whose length (500-650 bp) facilitated opportunities for more detailed phylogenetic  
7 analysis and the detection of subOTUs occurring in different habitats. To mitigate the possible  
8 problems arising from errors during demultiplexing the reads (wrong assignment of reads to  
9 samples), we used a set of long barcodes (10 nt) with minimal edit distance equal to 4 and did allow  
10 only one mismatch in a barcode. As the error probability in raw reads is close to  $1e-3$ , the  
11 probability of erroneous read assignment due to one tag mutating into another is  $1e-09$  under  
12 assumption of independent mutations. Thus, chimera formation might be the only mechanism  
13 leading to tag misidentification in our case and, as we employed three-step chimera removal  
14 procedure, it might be safely assumed that the number of misidentified tags in our data was  
15 negligible.

16

17 The flows were extracted from the .sff files, forward and reverse reads separately (sffinfo), then  
18 they were assigned to the samples basing on the MID sequences, trimmed to min. 500 and max. 650  
19 flows (trim.flows) and denoised with AmpliconNoise algorithms (shhh.flows and shhh.seqs).  
20 Primers and MIDs were removed from the denoised sequences, the reverse reads were reverse  
21 complemented (trim.seqs), and the reads set was dereplicated (unique.seqs). The forward and  
22 reverse read sets were pooled (cat) and the whole set was dereplicated again and aligned to the  
23 SILVA template alignment (align.seqs). Reads covering the desired region of the alignment (pos.  
24 6500-22500) were chosen (screen.seqs) and gap only and terminal gap-containing columns were  
25 removed from the alignment (filter.seqs). The set was dereplicated again and residual sequencing  
26 and PCR noise was removed with Single Linkage pre-clustering (pre.cluster, Huse *et al.*, 2010).

Chimera identification and removal was performed in three rounds: i) with UCHIME (chimera.uchime, Edgar *et al.*, 2011), ii) with PERSEUS (chimera.perseus, Quince *et al.*, 2011) and iii) with chimera slayer (chimera.slayer, Haas et al., 2011) using the SILVA gold template alignment from <http://www.mothur.org/w/images/f/f1/Silva.gold.bacteria.zip> (accessed on September 4, 2014).

To increase taxonomic resolution, full-length sequences (list.seqs, get.seqs) were used for classification with a naive Bayesian classifier (classify.seqs, Wang *et al.*, 2007) with the SILVA 119 template and taxonomy files ([http://www.mothur.org/w/images/2/27/Silva.nr\\_v119.tgz](http://www.mothur.org/w/images/2/27/Silva.nr_v119.tgz), accessed on September 4, 2014) at the bootstrap confidence level of 80%. Taxa assigned as 'unknown' and 'Chloroplast' were removed from the final set. Average linkage (UPGMA) algorithm was used to construct OTUs at the 0.03 dissimilarity level, and singletons as well as doubletons were removed from the data (remove.rare).

To ensure that OTUs frequencies in the subsampled dataset are close to the original ones, the final reads set was subsampled ten times to 2500 reads per sample (sub.sample), read names were mangled to reflect their coming from a particular subsample (regular expressions in the sed editor), subsamples were combined (cat), the whole set was dereplicated and used for distance matrix calculation (dist.seqs) and OTU construction via average neighbor clustering at 97% similarity level (cluster). A shared OTU table was constructed (make.shared) and the table averaged over the subsamples (i.e. for each OTU numbers of reads found in each subsample were summed and the sum was divided by ten) was calculated with a Perl script (average\_shared.perl). OTUs were classified using consensus approach with SILVA 119 taxonomic assignment (classify.otu).

Details are given below:

#Prerequisites: Mothur 1.32 installed under Linux environment (executable present in a directory listed in \$PATH is assumed) , Lookup\_Titanium.pat in a directory visible for mothur, SILVA files in a directory visible for mothur, bash shell, vi and sed editors, Perl 5, sff files, oligos files with samples assignment.

```
53  #Lines starting with # are commentaries, other lines are code to be copied to a terminal.
54  # x, x1, etc. denote a generic filename.
55  #In mothur commands the number of processors can (and should) be changed to be lower than the
56  number of accessible processors
57
58  #cd to the directory where sff files are stored
59
60  mkdir forward reverse
61
62  mothur
63
64  #For each sff file execute:
65  sff.info(sff=x.sff, flow=T)
66  quit()
67
68  cd forward
69
70  #For each flow file execute:
71  ln -s ../x.flow .
72
73  #Start mothur:
74  mothur
75
76  #For each flow file execute:
77  trim.flows(flow=x.flow, oligos=x_f.oligos, pdiffs=2, bdiffs=1, processors=6)
78  shhh.flows(file=x.flow.files, processors=18)
```

```

79 shhh.seqs(fasta=x.shhh.fasta, name=x.shhh.names, group=x.shhh.groups)
80 #Include files derived from all sffs
81 trim.seqs(fasta=x.shhh.shhh_seqs.fasta, name=x.shhh.shhh_seqs.names, oligos=x_f.oligos,
82 pdiffs=2, bdiffs=1, processors=4)
83 system(cat x.shhh.shhh_seqs.trim.fasta x1.shhh.shhh_seqs.trim.fasta x2.shhh.shhh_seqs.trim.fasta >
84 bacteria_f.shhh.shhh_seqs.trim.fasta)
85 system(cat x.shhh.shhh_seqs.trim.names x1.shhh.shhh_seqs.trim.names
86 x2.shhh.shhh_seqs.trim.names > bacteria_f.shhh.shhh_seqs.trim.names)
87 system(cat x.shhh.shhh_seqs.groups x1.shhh.shhh_seqs.groups x2.shhh.shhh_seqs.groups >
88 bacteria_f.shhh.shhh_seqs.groups)
89 quit()
90
91 cd ../reverse
92
93 #For each flow file execute:
94 ln -s ../x.flow .
95
96 #Start mothur:
97 mothur
98
99 #For each flow file execute:
100 trim.flows(flow=x.flow, oligos=x_r.oligos, pdiffs=2, bdiffs=1, processors=6)
101 shhh.flows(file=x.flow.files, processors=18)
102 shhh.seqs(fasta=x.shhh.fasta, name=x.shhh.names, group=x.shhh.groups, processors=1)
103 #Include files derived from all sffs
104 trim.seqs(fasta=x.shhh.shhh_seqs.fasta, name=x.shhh.shhh_seqs.names, oligos=x_f.oligos,

```

```
105 pdiffs=2, bdiffs=1, reverse=T, processors=4)
106 system(cat x.shhh.shhh_seqs.trim.fasta x1.shhh.shhh_seqs.trim.fasta x2.shhh.shhh_seqs.trim.fasta >
107 bacteria_r.shhh.shhh_seqs.trim.fasta)
108 system(cat x.shhh.shhh_seqs.trim.names x1.shhh.shhh_seqs.trim.names
109 x2.shhh.shhh_seqs.trim.names > bacteria_r.shhh.shhh_seqs.trim.names)
110 system(cat x.shhh.shhh_seqs.groups x1.shhh.shhh_seqs.groups x2.shhh.shhh_seqs.groups >
111 bacteria_r.shhh.shhh_seqs.groups)
112 quit()
113
114 cd ..
115
116 cat forward/bacteria_f.shhh.shhh_seqs.fasta reverse/bacteria_r.shhh.shhh_seqs.fasta > bacteria.fasta
117 cat forward/bacteria_f.shhh.shhh_seqs.names reverse/bacteria_r.shhh.shhh_seqs.names >
118 bacteria.names
119 cat forward/bacteria_f.shhh.shhh_seqs.groups reverse/bacteria_r.shhh.shhh_seqs.groups >
120 bacteria.groups
121
122 mothur
123
124 unique.seqs(fasta=bacteria.fasta, name=bacteria.names)
125 align.seqs(fasta=current, reference=silva.bacteria.fasta, processors=16)
126 remove.seqs(fasta=current, name=current, group=bacteria.groups, accnos=current)
127 screen.seqs(fasta=current, name=current, group=current, start=6500, end=22500)
128 filter.seqs(fasta=current, vertical=T, trump=.)
129 unique.seqs(fasta=current, name=current)
130 pre.cluster(fasta=current, name=current, group=current)
```

```

131 chimera.uchime(fasta=current, name=current, group=current, reference=groups)
132 remove.seqs(fasta=current, name=current, group=current, accnos=current)
133 chimera.perseus(fasta=current, name=current, group=current)
134 remove.seqs(fasta=current, name=current, group=current, accnos=current)
135 chimera.slayer(fasta=current, name=current, group=current, reference=silva.gold.fasta)
136 remove.seqs(fasta=current, name=current, group=current, accnos=current)
137 list.seqs(fasta=current)
138 get.seqs(fasta=bacteria.fasta, accnos=current) #get full length seqs for classification
139 classify.seqs(fasta=current, name=current, group=current, reference=silva.bacteria.fasta,
140 taxonomy=silva.bacteria.tax, cutoff=80)
141 remove.lineage(fasta=bacteria.unique.pick.good.filter.unique.precluster.pick.pick.pick.fasta,
142 name=current, group=current, taxonomy=current, taxon=Bacteria;Cyanobacteria;Chloroplast;)
143 remove.lineage(fasta=current, name=current, group=current, taxonomy=current, taxon=unknown;)
144 dist.seqs(fasta=current, cutoff=0.10, processors=16)
145 cluster(column=current, name=current)
146 remove.rare(list=current, label=0.03, nseqs=2)
147 list.seqs(list=current)
148 get.seqs(fasta=current, name=current, group=current, accnos=current) #get seqs set without
149 singletons and doubletons
150 quit()
151
152 mv bacteria.unique.pick.good.filter.unique.precluster.pick.pick.pick.pick.pick.fasta
153 bacteria.final.fasta
154
155 mv bacteria.unique.pick.good.filter.unique.precluster.pick.pick.pick.pick.names
156 bacteria.final.names

```

```

157
158 mv bacteria.pick.good.pick.pick.pick.pick.pick.groups bacteria.final.groups
159
160 mv bacteria.unique.pick.good.filter.unique.precluster.pick.pick.pick.pick.pick.0.03.an.pick.list
161 bacteria.final.an.list
162
163 #The procedure below was devised to mitigate the effect of single subsampling, namely possibility
164 of OTU frequencies being far off the real ones (meaning the frequencies in the whole dataset). Ten
165 subsamples are generated, read names are mangled to reflect their coming from a particular
166 subsample, the resulting set is dereplicated and OTUs are constructed as above. Shared OTU table
167 is then constructed and averaged over the subsamples (i.e. numbers of reads coming from a given
168 OTU in each subsample are summed and the result is divided by the number of subsamples). The
169 reads are classified and the results are averaged analogically, but at taxa levels instead of OTUs.
170 There is a possibility of bootstrapping in some mothur commands, such as unifrac.(un)weighted,
171 summary.single or dist.shared. Its was used here.
172
173 for f in 1 2 3 4 5 6 7 8 9 10; do mothur „#sub.sample(fasta=bacteria.final.fasta,
174 name=bacteria.final.names, group=bacteria.final.groups, pergroup=T, size=2500);”; cat
175 bacteria.final.subsample.fasta | sed „s/>/>$f\_” >> bacteria.bootstrap.fasta; cat
176 bacteria.final.subsample.names | sed „s/^/$f\_” | sed „s/t/t$f\_” | sed „s/,\/, $f\_/g” >>
177 bacteria.bootstrap.names; cat bacteria.final.subsample.groups | sed „s/^/$f\_” | sed „s/_/$f” >>
178 bacteria.bootstrap.groups; done
179
180 mothur
181 unique.seqs(fasta=bacteria.bootstrap.fasta, name=bacteria.bootstrap.names)
182 list.seqs(fasta=bacteria.bootstrap.unique.fasta)

```



```

183 dist.seqs(fasta=current, cutoff=0.10, processors=16)
184 cluster(column=current, name=current)
185 make.shared(list=current, group=bacteria.bootstrap.groups, label=0.03) #shared OTU table for
186 averaging
187 make.shared(list=bacteria.final.an.list, group=bacteria.final.groups, label=0.03) #shared OTU table
188 for diversity estimations and generation of community distance matrices
189 dist.shared(shared=current, calc=braycurtis-morisitahorn, subsample=2500, iters=100)
190 summary.single(shared=current, calc=sobs-chao-ace-shannon-shannoneven, subsample=2500,
191 iters=100)
192 clearcut(fasta=bacteria.final.fasta, DNA=T, kimura=T)
193 unifracs.weighted(tree=current, name=bacteria.final.names, group=bacteria.final.groups,
194 subsample=2500, distance=lt, processors=16)
195 quit()
196
197 extract_full_length_seqs.perl -l bacteria.bootstrap.unique.accnos -f bacteria.fasta >
198 bacteria.bootstrap.unique.fulllength.fasta #the script fetches sequences from a fasta file whose names
199 are those from the accnos file with subsample number dropped, sequences from the fasta file are
200 printed with names coming from the accnos file
201
202 mothur
203 classify.seqs(fasta=bacteria.bootstrap.unique.fulllength.fasta,
204 name=bacteria.bootstrap.unique.names, group=bacteria.bootstrap.groups,
205 reference=silva.bacteria.ng.fasta, taxonomy=silva.bacteria.tax, cutoff=80, probs=F, processors=16)
206 #no bootstrap probabilities, they preclude OTUs classification with classify.otu
207 classify.otu(list=bacteria.bootstrap.unique.an.list,
208 taxonomy=bacteria.bootstrap.unique.fulllength.wang.taxonomy,

```

```

209 name=bacteria.bootstrap.unique.names, cutoff=80)
210 quit()
211
212 average_shared.perl bacteria.bootstrap.unique.an.shared > bacteria.bootstrap.unique.an.avg.shared
213
214 average_tax.summary.perl -f bacteria.bootstrap.unique.fulllength.wang.tax.summary -n 10 >
215 bacteria.bootstrap.unique.fulllength.avg.tax.summary.csv
216
217
218 #For vegan-based analyses the shared OTUs file was manually edited in vi to remove a redundant
219 tabulator at the end of the header line and was imported to R
220
221 R
222 bacteria.community <- read.table(„bacteria.bootstrap.unique.an.avg.shared”, header=T, sep=”\t”,
223 dec=”.”)
224 rownames(bacteria.community) <- bacteria.community$Group
225 bacteria.community$Group <- NULL
226 bacteria.community$label <- NULL
227 bacteria.community$numOtus <- NULL
228
229 #Construction for subOTUs for 50 most abundant OTUs
230 for f in {1..50}; do get_otu_reads_accnos.perl bacteria.bootstrap.unique.an.list 0.03 $f >
231 otu$f\accnos; mothur „#get.seqs(fasta=bacteria.bootstrap.unique.fasta,
232 name=bacteria.bootstrap.unique.names, group=bacteria.bootstrap.groups, accnos=otu$f\accnos);”;
233 mv bacteria.bootstrap.unique.pick.fasta otu$f\fasta; mv bacteria.bootstrap.unique.pick.names
234 otu$f\names; mv bacteria.bootstrap.pick.groups otu$f\groups; mothur

```

```

235  ,#dist.seqs(fasta=otu$f\fasta, cutoff=0.10, processors=4); cluster(column=otu$f\dist,
236  name=otu$f\names); make.shared(list=otu$f\an.list, group=otu$f\groups, label=0.01);
237  get.oturep(list=otu$f\an.list, column=otu$f\dist, name=otu$f\names, fasta=otu$f\fasta,
238  label=0.01, method=distance, weighted=T); clearcut(fasta=otu$f\an.0.01.rep.fasta, DNA=T,
239  kimura=T);"; cat otu$f\an.0.01.rep.tre | sed „s/Otu/subOtu/g” > otu$f\an.0.01.rep.mod.tre; cat
240  otu$f\an.shared | sed „s/Otu/subOtu/g” > otu$f\an.mod.shared; done
241
242  #Trees generated by version of clearcut incorporated into mothur are sometimes not conforming to
243  the standard and need to be manually edited to be correctly read by phyloseq's import_mothur
244  function. The problem lies in an unnecessary pair of parentheses, where the closing one directly
245  precedes a comma. This pair should be removed.
246  #Sample data file should be prepared as a tab-separated file. The file should include site and season
247  for each sample.
248
249  R
250  library(phyloseq)
251  sdata ← read.table(„sample_data.csv”, header=T, sep=”\t”);
252  sdata$site ← factor(sdata$site, levels=c('freshwater','mixing_zone', 'brackish'))
253  sdata$season ← factor(sdata$season, levels=c('spring', 'summer', 'autumn', 'winter'))
254  #For each OTU execute
255  otux ← import_mothur(mothur_shared_file=”otux.an.mod.shared”,
256  mothur_tree_file=”otux.an.0.01.rep.tre”, cutoff=0.01)
257  sample_data(otux) ← sample_data(sdata)
258  pdf(file=”otux.pdf”)
259  print( plot_tree(otux, shape=”season”, color=”site”, size=”abundance”, label.tips=”taxa_names”,
260  title=”OTUx”) )

```

```

261 dev.off()
262
263 #The pdf files may be collated later, or printing may be performed within a 'for' loop with
264 pdf(file="...", onefile=T)
265
266 References
267     Edgar R.C., Haas B.J., Clemente J.C., Quince C., and Knight R. (2011) UCHIME improves
268 sensitivity and speed of chimera detection. Bioinformatics 27: 2194-2200.
269     Haas B.J., Gevers D., Earl A.M., Feldgarden M., Ward D.V., Gannoukos G., et al. (2011)
270 Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR
271 amplicons. Genome Res 21: 494-504.
272     Huse S.M., Welch D.M., Morrison H.G., and Sogin M.L. (2010) Ironing out the wrinkles in the
273 rare biosphere through improved OTU clustering. Environ Microbiol 12: 1889-98.
274     Quince C., Lanzen A., Davenport R.J., and Turnbaugh P.J. (2011) Removing noise from
275 pyrosequenced amplicons. BMC Bioinformatics 12: 38.
276     Wang Q., Garrity G.M., Tiedje J.M., and Cole J.R. (2007) Naive Bayesian classifier for rapid
277 assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 63:
278 5261-5267.

```

Table S2. Significance of clusters separation for unweighted UniFrac distance matrix. Significant comparisons were marked with boldface font.

Comparison	AMOVA		ANOSIM		Bonferroni corrected significance threshold
	F	P	R	P	
HEL-KIEZ	<b>6.985</b>	<b>1e-06</b>	<b>0.982</b>	<b>&lt;1e-06</b>	1.7e-02
HEL-UJS	<b>4.145</b>	<b>1e-06</b>	<b>0.729</b>	<b>&lt;1e-06</b>	
UJS-KIEZ	1.570	3.1e-02	0.162	3.4e-02	
spring-summer	<b>3.382</b>	<b>1.4e-04</b>	<b>0.551</b>	<b>6.2e-05</b>	8.3e-03
spring-autumn	<b>3.097</b>	<b>3.6e-04</b>	<b>0.527</b>	<b>3.1e-04</b>	
spring winter	2.032	1.3e-02	0.324	1.1e-02	
summer-autumn	2.469	9.9e-03	<b>0.341</b>	<b>7.9e-03</b>	
summer-winter	<b>3.424</b>	<b>3.8e-05</b>	<b>0.539</b>	<b>2.4e-04</b>	
autumn-winter	<b>2.418</b>	<b>6.6e-03</b>	<b>0.385</b>	<b>4.9e-03</b>	